Norwegian University of Life Sciences
School of Economics and Business

Philosophiae Doctor (PhD)
Thesis 2022:4

Drivers of deforestation and incentives for REDD+

Drivere for avskoging og insentiver for REDD+

Julia del Carmen Naime Sánchez Henkel

Drivers of deforestation and incentives for REDD+

Drivere for avskoging og insentiver for REDD+

Philosophiae Doctor (PhD) Thesis

Julia del Carmen Naime Sánchez Henkel

Norwegian University of Life Sciences
School of Economics and Business

Ås (2022)

Thesis number 2022:4
ISSN 1894-6402
ISBN 978-82-575-1881-3

Supervisors and Evaluation Committee

PhD supervisors

Main supervisor: Arild Angelsen, Professor
School of Economics and Business
Norwegian School of Life Sciences (NMBU)
Ås, Norway

Co-supervisor: Frode Alfnes, Professor
School of Economics and Business
Norwegian School of Life Sciences (NMBU)
Ås, Norway

Evaluation committee

Unai Pascual, Ikerbasque Professor
Basque Center for Climate Change
Bilbao, Spain

Torfinn Harding, Professor
Department of Economics and Finance
University of Stavanger Business School
Stavanger, Norway

Ståle Navrud, Professor
School of Economics and Business
Norwegian School of Life Sciences (NMBU)
Ås, Norway

Dedication

To my family: Mom, Dad, Nemer, Yamel, Ivonne, and Benito. One does not get to choose the family one is born into, so I was truly lucky to have landed in yours.

This research is part of CIFOR's Global Comparative Study on REDD+ (www.cifor.org/gcs). The funding partners that have supported this research include the Norwegian Agency for Development Cooperation (Norad, Grant No. QZA-16/0110 & No. QZA-21/0124), International Climate Initiative (IKI) of the German Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU, Grant No. 20_III_108), and the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA) with financial support from the CGIAR Fund Donors.

Acknowledgements

Getting to the finish line of a PhD is really a collective effort, so there are many people to thank. The deepest thanks go to Arild Angelsen, for offering fantastic support and patience, encouragement and advice in the last four years, and for all the opportunities and trust given to me as a PhD student. Thank you for creating such a pleasant, friendly, supportive and stimulating work environment, which fosters creativity, serenity, and trust, and for always offering a smile or a joke in meetings while also helping me solve the hardest problems and questions. It has been a pleasure and a privilege, and I know I would not have had it better or been happier doing the PhD anywhere else. So, hjertelig takk for making of the PhD such a nice, interesting, and enriching journey. Words are not lacking and I could go on, but I do need space to thank the others.

Thanks to my coauthors, A. Angelsen, E.O. Sills, A. Molina, A.E. Duchelle, C. Martius for their extremely useful feedback, encouragement, and rich discussions. Your thoughts and comments were absolutely necessary to improve and maintain my efforts. Thanks to D. Rodríguez-Ward, V. Selviana, C. Carrilho, G. Diaz for the relevant feedback and data collection efforts. Special thanks to Amy Duchelle also for all her support and encouragement to PhD students.

Thanks to CIFOR for being such a nice network of researchers, for facilitating fieldwork, for explicitly and implicitly teaching good research ethics, and for all their support to PhD students. Thanks to Techy and Abelardo for providing (literally and figuratively) the warmest home in Pucallpa. Thanks to all my enumerators, in particular Medardo Miranda, Alex Talaverano and Pilar Paredes. Thanks Dawn and Amy for training me and explaining the basics of fieldwork.

Thanks to my co-supervisor Frode Alfnes, for the extremely useful feedback and tips, as well as Stein Holden, Ragnar Øygard, Eirik Romstad, for providing sharp and timely advice to improve the work presented here.

Thanks to Arild, Niki De Sy, and C. Martius for offering me this PhD opportunity, and T. Pham for the continued support.

Thanks to all the Shipibo-Konibo families in Ucayali (Peru), who had the patience to answer all our questions. To you I am indebted.

Thanks to NMBU and NMBU Handelshøyskolen and its faculty, for providing a comfortable work and office environment, good and exciting courses, amicable faculty members and colleagues, great administrative support (with Kirsti, Reidun, Hanne, Berit), unlimited coffee, and a massage chair.

Thanks to my lovely office mates, Viktorija (the golden age of office sharing, it will never be the same! Thanks for always being there, you were definitely the Sam to my Frodo), and Diana (we can do it!). Thanks also to Sodoc's network and activities.

Thanks to all my friends who have closely accompanied me in the process: Viktorija, Lulú, Øyvind, Sergio, Sara, Stine, Olga, Kat, April, Fede, Marius and to those who for the sake of not making this list unreasonably long I cannot mention. You have been a central source of smiles, laughs, and basic wellbeing in the last four years, and in fact, I also learned a lot from you and your wonderful personalities, so thank you for that too.

Thanks to my friends back home in Mexico and those spread throughout the world, who despite the distance have been there with me (Santi, Mike, Nat, Luci, Marta, Vans, niñas).

Thanks to Yamel and Kristian, for being unfailingly nice and welcoming when I visited you in Bergen. The pandemic has been hard on many of us, I have been lucky to have some family close.

Thanks also to Patricia Balvanera, for all the initial encouragement and support, and contagious passion and dedication. In a similar vein, thanks to Julia Carabias. Thanks also to my teachers of the MSc. Environmental Economics and Climate Change at the LSE for designing and teaching a really nice curriculum that ultimately led me to this PhD.

And finally, big thanks to my family: Dad, Mom, Nemer, Yamel, Ivonne and Benito. You have always been my greatest source of support and encouragement. I admire and cherish you all.

Table of Contents

List of papers

Paper I. Enforcement and inequality in collective PES: efficiency, effectiveness and equity implications. *Global Environmental Change,* revise and re-submit.
Authors: J. Naime, A. Angelsen, A. Molina-Garzón, C.D. Carrilho, V. Selviana, G. Demarchi, A.E. Duchelle and C. Martius

Paper II. Will peer punishment protect tropical forests? Multi-country evidence from framed field experiments
Authors: J. Naime, A. Angelsen

Paper III. Land use and livelihood impacts of two collective incentive-based conservation interventions in Ucayali, Peru
Authors: J. Naime, A. Angelsen, E.O. Sills, D. Rodriguez-Ward

Paper IV. Re-examining the macroeconomic drivers of deforestation
Author: J. Naime

Abstract

Tropical deforestation is a global environmental and development problem. Identifying policies that effectively reduce deforestation while improving rural livelihoods is essential to meet our climate and sustainable development goals. The thesis has two main objectives: first, to evaluate the potential of collective Payments for Ecosystem Services (PES) to deliver on conservation and development outcomes under different contexts, and second, to improve our understanding of the causes of deforestation at the national level.

Paper I and II investigate the central question of how to solve the free rider problem of collective PES, using data from a framed field experiment in Brazil, Indonesia and Peru. They examine three strategies to mitigate the free-rider problem: (i) increase in public monitoring of individual actions, (ii) peer-to-peer, community sanctions, and (iii) external, government sanctions. Overall, public monitoring and both types of sanctions increase policy effectiveness. Government sanctions are the most effective to reduce deforestation; community or peer-to-peer sanctions also reduce deforestation but can create trade-offs in terms of loss of local income. There are also important cross-country differences in policy impact. Increased public monitoring does not improve the performance of collective PES in Brazil, while in Indonesia peer-to-peer sanctions are much more frequent. In Indonesia, the existence of inequality in wealth reduces the performance of collective PES, while it has negligeable impacts in Peru and Brazil. In general, individuals who contribute more to conservation outcomes are also the ones who contribute more to the enforcement of conservation norms by sanctioning free-riders.

Paper III presents a household-level impact evaluation of two collective PES schemes in Ucayali, Peru. The first is a local conservation project led by an NGO while the second is the Peruvian's government National Forest Conservation Program (NFCP). The paper examines land use, income, and wellbeing outcomes. The projects have not improved local incomes because of a slow and delayed implementation. The delayed and slow implementation have not, in turn, negatively affected forest or conservation outcomes. The study brings forward the importance of considering households' subjective evaluations to identify possible 'intangible' wellbeing impacts of conservation projects. Taken together, the results of Paper I, II and III point to the importance of having good forest monitoring to improve conservation, development, and wellbeing outcomes of collective PES.

Paper IV addresses the second objective of the thesis by examining how dual economy growth models can help understand patterns of deforestation across countries. The paper develops a theoretical framework to disentangle the immediate drivers of deforestation from the indirect drivers. The results are consistent with economic predictions: competing land use value between forest and agriculture are a major immediate driver of deforestation. There is suggestive evidence that openness to trade can, indirectly, reduce deforestation levels by decreasing the relative return of agricultural land.

Norsk sammendrag

Tropisk avskoging er et globalt miljø- og utviklingsproblem. Redusert avskoging og forbedring av levebrødet på landsbygda er avgjørende for å oppfylle mål for reduserte klimaendringer og bærekraftig utvikling. Avhandlingen har to hovedmål: for det første å evaluere effekten av politikk som tar sikte på å redusere avskoging og forbedre levebrødet på landsbygda, med et fokus på kollektive betalinger for økosystemtjenester (PES), og for det andre å forbedre vår forståelse av årsakene til avskoging.

Artikkel I og II undersøker det sentrale spørsmålet om hvordan man kan løse gratispassasjer problemer ved kollektive PES, ved å bruke data fra et økonomisk felteksperiment i Brasil, Indonesia og Peru. Artiklene undersøker tre strategier for å redusere gratispassasjerproblemet: (i) økt offentlig overvåking av enkeltpersoners handlinger, (ii) interne sanksjoner fra lokalsamfunnet, og (iii) eksterne sanksjoner fra myndighetene. Offentlig overvåking og begge typer sanksjoner øker effektiviteten av PES. Offentlige sanksjoner er det mest effektive virkemiddelet for å redusere avskogingen. Sanksjoner fra lokalsamfunnet reduserer også avskogingen, men kan skape målkonflikter ved at de også medfører redusert lokal inntekt. Det er også viktige forskjeller i politikkeffekter på tvers av land. Økt overvåking forbedrer ikke effektiviteten til kollektiv PES i Brasil, mens sanksjoner fra lokalsamfunnet er mye hyppigere brukt eksperimentene i Indonesia. Generelt er deltagerne som bidrar mest til skogvern også de som bidrar mest til håndhevelsen av vernenormene gjennom å straffe gratispassasjerer. I Indonesia reduserer eksistensen av ulikhet effektiviteten av kollektiv PES. Samlet sett understreker resultatene viktigheten av god overvåking og også individuelle sanksjoner for å verne tropisk skog.

Artikkel III presenterer en evaluering på husholdningsnivå av to kollektive PES systemer i Ucayali, Peru. Det første er et lokalt REDD+-prosjekt ledet av en NGO, mens det andre er den peruanske regjeringens National Forest Conservation Program (NFCP). Artikkelen undersøker effekten på arealbruk, inntekt og velstand. Prosjektene har ikke økt lokale inntekter på grunn av en langsom og forsinket implementering. Likevel finner vi ikke grunnlag for å konkludere at forsinkelsene har påvirket verneresultatene negativt. Artikkelen peker på viktigheten av å studere lokalsamfunnets subjektive vurderinger for å identifisere mulige immaterielle effekter og lokalsamfunnenes støtte og involvering i verneprosjekter. Samlet sett peker resultatene av artikler I, II og III på viktigheten av å ha gode overvåkings for å forbedre bevaring, utvikling og velvære i kollektive PES.

Artikkel IV undersøker hvordan vekstmodeller for duale økonomier og teorier om strukturelle endringer kan bidra til å forstå underliggende mønstre for avskoging på tvers av land. Artikkelen utvikler et teoretisk rammeverk for å analysere driverne bak avskoging og skiller mellom direkte og indirekte drivere. Resultatene samsvarer med prediksjonene fra økonomisk teori: forskjeller i arealbruksverdien mellom skog og jordbruk som er den viktigste direkte driveren for avskoging. Åpenhet for handel kan, indirekte, redusere avskogingen gjennom redusert avkastning fra jordbruksland.

1 Introduction

"*Vue du dehors, la forêt amazonienne semble un amas de bulles figées, un entassement vertical de boursouflures vertes. Mais quand on crève la pellicule et qu'on passe au-dedans, tout change : vue de de l'intérieur, cette masse confuse devient un univers monumental.* » Claude Lévi-Strauss, Tristes Tropiques

Seen from the outside, the Amazon rainforest looks like a mass of inert bubbles, a vertical pile of green blisters. But when you break the film and dare to go inside, everything changes: seen from the inside, this confused mass becomes a monumental universe.

1.1 Setting the stage

Ending deforestation is a climate change, development, and international cooperation challenge. To address deforestation there is a need to: (i) adequately understand the causes of deforestation and land-use change and (ii) design and implement effective, efficient and equitable policies that tackle deforestation while also improving local livelihoods. The following thesis aims to contribute to these two research needs, drawing from development, behavioural and experimental economics theories and methods.

1.1.1 Motivation

Forests provide multiple and essential ecosystem services to humans. They play an essential role in shaping the climate conditions that have allowed human species to flourish (Bonan, 2008; Steffen et al., 2015), they are essential elements of the global water and carbon cycle (Sheil, 2018; Harris et al., 2021), and they harbour unique and vast amounts of the world's terrestrial biodiversity (Raven et al., 2020). Forests are also a primary source of raw products such as timber, medicines, and food, and they play an important role in regulating and preventing the spread of diseases, such as malaria and ebola (Olivero et al., 2017; Chaves et al., 2020; Gibb et al., 2020).

Forests are also a global net carbon sink, which is important for climate change mitigation efforts (Harris et al., 2021). It is estimated that the protection and restoration of the world's forest can deliver up to 37% of emissions mitigation needed to stay within 2ºC of warming (Griscom et al., 2017). Thus, forests are at the forefront for meeting our climate change and biodiversity conservation goals.

Yet, forests – particularly in tropical regions – are facing considerable threats. Although in the last three decades we have gained global forest cover due to afforestation and reforestation in temperate zones (Song et al., 2018), current trends indicate that tropical deforestation has increased in recent years, from 7.5 million hectares (Mha) lost annually in 2001 to 18.9 Mha lost in 2017 (Hansen et al., 2013; Angelsen et al., 2018). Deforestation is the most important source of greenhouse gas emissions from the land-use sector, and sums to approximately 10-12% of global emissions (IPCC 2019).

The vast majority of global deforestation is concentrated in the tropics, in primary forest cover (Song et al., 2018; Turubanova et al., 2018). During the 2001 to 2015 period, the main drivers of global forest loss were: agricultural commodities driving 27% of forest loss, forestry commodities (e.g., plantations) driving 26%, and shifting agriculture and wildfires were responsible for 24% and 23% of the change, respectively (Curtis et al., 2018). This human lead land-use change is threatening multiple species and ecosystem services and is further exacerbated by current climate warming (particularly wildfires).

Concerningly, continued deforestation also threatens the carbon sink potential of forests such as the Amazon (Mitchard, 2018; Gatti et al., 2021), as well as long-term forest quality and biodiversity (Edwards et al., 2019). Most of the remaining standing, old-growth tropical forests host, in fact, 'irrecoverable' carbon (Noon et al., 2021). The global social costs of current and future deforestation are substantial, considering that future deforestation could release up to 169 giga-tons of CO_2 in the 2016-2050 period (Busch and Engelmann, 2017), and that recent estimations of the social cost of carbon (SCC) vary between \$31.2 and \$197.4 (in 2010 USD) per ton of CO_2 (Nordhaus et al. 2017).

At the same time, an estimated 1.6. billion people living near forests struggle to meet their basic needs without further degrading or deteriorating the forest (Newton et al., 2016; Newton et al., 2020). Indeed, many poor rural households are dependent on forest use and its transformation (Angelsen et al., 2014). Rural households are of central importance to reduce extreme poverty rates (Castañeda et al., 2018). Hence, equitable approaches to reduce deforestation require addressing the development needs of the rural poor.

Although reducing deforestation has been portrayed as the 'low-hanging fruit' of climate mitigation because of relatively low abatement costs as compared to other sectors (Angelsen and Wertz-Kanounnikoff, 2008; Busch and Engelmann, 2017), reducing global deforestation has not been an easy task. The poverty alleviation and development needs of rural households dependent on forests increase the political costs of forest conservation policies. Public and private efforts have fallen short in the ambitious goals of halving deforestation by 2020 and stopping it by 2030 (NYDF, 2021).

The lack of success in reducing deforestation can be partly explained by lack of global cooperation. As a global climate change mitigation action, reducing deforestation suffers from the basic cooperation problem innate to public goods: the benefit is global, but the costs are often incurred at local and regional scales. Recent international cooperation regarding forest conservation has been mostly embodied in the initiative Reducing Emissions from Deforestation and Degradation (REDD+). Starting in 2007 and initially conceived as a large international Payment for Ecosystem Services (PES) between nations, REDD+ has suffered multiple conceptual and implementation changes (Angelsen et al., 2017). Institutional, political and economic circumstances inhibit the implementation of relevant policies at the international and national level. The lack of a global carbon market integrating REDD+ credits as an offsetting option has made funding much less than initially envisioned. As a result, the relationship between donors and recipients has been mostly transformed to 'result-based aid' (Angelsen, 2017).

Reflecting the demands of donors, recipients, and civil society groups, REDD+ thus evolved to include more objectives than forest conservation, such as improved livelihoods, biodiversity and adaptation. The funding from international cooperation has been too little and too slow (Atmadja et al., 2018). The local (political and economic) burden of reducing deforestation is often seen by developing countries as being too high, and the compensation offered by developed countries not high enough. There is thus a need to design, implement, and enforce policies as effectively and efficiently as possible, without negatively affecting the livelihoods of the poor.

1.1.2 Objectives

The thesis consists of this introductory chapter and four research articles (Paper I to Paper IV). It has two main research objectives. First, to evaluate the performance of policies and policy mixes aiming to reduce deforestation and improve rural livelihoods, with a specific focus on collective Payments for Ecosystem Services (PES). Collective PES are positive incentives that reward forest users conditional on collective conservation performance, and they have increasingly caught scholarly attention (Hayes et al., 2019). Second, it aims to improve our understanding of the causes of deforestation: what explains changes in forest cover and deforestation as economies grow? Paper I to III focus on evaluating the performance of collective PES under different conditions, while the last paper addresses the question of national-level drivers of deforestation.

Paper I, “Enforcement and inequality in collective PES: efficiency, effectiveness and equity implications” addresses the central question of how to solve the free rider problem of collective PES. In collective PES individual participants have an incentive to free ride on others’ conservation behaviour because the rewards are based on group rather than individual performance. It examines three free-riding mitigation strategies: (i) public monitoring of individual deforestation, (ii) peer punishment, and (iii) external enforcement. Using a framed field experiment in Brazil, Indonesia and Peru, it compares their relative performance in terms of effectiveness, efficiency and equity (3Es), and whether the existence of inequality in wealth moderates their impacts.

Paper II, “Will peer punishment protect tropical forests? Multi-country evidence from framed field experiments” delves deeper into the patterns and dynamics of peer punishment, one of the free-riding mitigation strategies addressed in Paper I. The main objective of the paper is to examine the relationship between first and second order cooperation, and to identify whether peer punishment effectively reduces deforestation in the country sites. This paper uses data from the same framed field experiment conducted in Brazil, Indonesia and Peru, and analysed in Paper I.

Paper III, “Land use and livelihood impacts of two collective incentive-based conservation interventions in Ucayali, Peru” is an impact evaluation of two collective PES as they take place on the ground. The projects are implemented in Peru, the country with the second largest area of the Amazon rainforest. One of the projects is a government financed collective PES, while the second project is a collective PES implemented by a local NGO. The paper examines impacts on land-use, livelihood, and wellbeing variables.

Paper IV, “Re-examining the macroeconomic drivers of deforestation” examines the important question of drivers of deforestation at the national level. It examines how theories of development economics can help understand patterns of deforestation across countries. The paper utilizes data from international sources such as the World Bank’s World Development Indicators, and the Food and Agriculture Organization (FAO) data bases to construct indicators of structural change in national economies and examine their links to deforestation rates.

The research and doctoral project was conducted in close collaboration with the Center for International Forestry Research (CIFOR), as part of their research project the Global Comparative Study on REDD+ (GCS REDD+). Initiated in 2010 and currently in its fourth phase, the goal of the GCS REDD+ is to evaluate the progress and impact of international REDD+ policies on the ground. One of the most important components and goals of CIFOR's GCS REDD+ project has been to evaluate impacts of 23 sub-national REDD+ projects in six countries (Peru, Brazil, Cameroon, Tanzania, Indonesia, and Vietnam). The multi-country comparative research project aims to monitor and evaluate implementation covering up to 150 communities (with REDD+ interventions or not).

The rest of this chapter is structured as follows. Section 1.2 presents the main theories and literature used to address the research objectives of each paper. Section 1.3 presents an overview of the methods and study sites. Section 1.4 describes and discusses the data and the data collection process, followed by the presentation of main findings, and contribution in Section 1.5. Section 1.6 finalizes the chapter with concluding remarks.

1.2 Theoretical background

The following section presents how the problem of deforestation is a global externality and thus a relevant object of study of environmental and public economics (section 1.2.1). It then presents a brief overview of current evidence of policies to reduce deforestation (section 1.2.2), before delving into the specific theories and research questions guiding each paper (section 1.2.3 to section 1.2.5).

1.2.1 Deforestation as a global externality and market failure

Externalities occur when the production or consumption decision of an agent affect the utility of another agent. The concept of externalities was introduced by Pigou (1920), who separated between economic (or private) and social (or public) welfare. When there is an externality in an unregulated market, there is sub-optimal production of the good or service in question, as all the social, public benefits or loses are not correctly accounted for. Externalities can be positive or negative. When the externalities are positive the good is underproduced, while when externalities are negative, the good in question is overproduced.

As discussed in section 1.1, tropical deforestation is a global negative externality because forest loss entails costs to society (including future generations) in terms of biodiversity loss, climate change, and loss of local ecosystem services. These costs are not considered by the agents of deforestation (e.g., a firm, a smallholder, or a local government), who – assuming selfish decision making – solely base their decisions on the private costs and benefits of deforestation. Reducing deforestation often requires government intervention given that some of the public benefits of forests, such as climate regulation, are shared among the 7.9 billion people on this planet, making it hard for individuals to coordinate on its provision.

A social planner interested in maximizing societal welfare and internalize the externality of deforestation has three main strategies:

i) Coasian solution: create property rights so that the polluter either receives the right to pollute, or the victim receives the right to claim a compensation for the damages of pollution. If there are no transaction costs between the 'polluter' and the affected, bargaining between the agents can lead to an efficient solution and internalization of the externality (Coase, 1960).

ii) Market solutions: price-based instruments, such as taxes or subsidies can lead to efficient outcomes when they are equal to the marginal damage. The Pigouvian tax rate should be the size of the externality to reduce the quantity of deforestation to the social optimal amount.

iii) Command and control: quantity-based instruments. In such cases, a regulatory framework and enforcement is introduced in order to reduce the quantity of deforestation to the social optimal.

The optimal policy choice depends on who are the involved agents, their bargaining power, the existence of uncertainties regarding the size of the externality, the slope of the marginal cost and benefit curves, and the transaction costs and equity considerations.

1.2.2 Overview of policies to reduce deforestation

Policies to reduce deforestation can be broadly classified into three categories aligned to the three main strategies described in section 1.2.1 (Börner et al., 2020): enabling policies, disincentives, and incentives.

Enabling policies. These policies include property rights and land tenure, increasing tenure security, or decentralization and devolution of rights to local communities (Börner et al., 2020). There is mixed evidence regarding the impact of such policies on deforestation. For instance, for the effect of tenure clarification on deforestation, studies of land titling initiative in Brazil find no or negative effects on deforestation (BenYishay et al., 2017; Probst et al., 2020), but positive effects in Peru (Blackman et al., 2017). Decentralization of forest management can also help reduce deforestation, because local stakeholders have better knowledge about local contexts and can help monitoring (e.g., Oldekop et al., 2019), but it does not necessarily lead to improved forest management (Wright et al., 2016; Jagger et al., 2018).

Disincentive-based policies. These are 'sticks' that aim to increase the costs of deforestation and forest degrading activities. This set of policies includes the establishment of protected areas (PA), increased environmental regulation and enforcement, or taxes on environmental harmful products. While many PAs work as 'paper parks' (Blackman et al., 2015) because of low level of enforcement and additionality[1], and there is a strong bias in setting PAs in areas with low opportunity costs (i.e., setting them 'high and far') (Joppa and Pfaff, 2010), they in general decrease deforestation (Joppa and Pfaff, 2011). Besides PAs, prominent cases of relatively successful disincentives policies are Indonesia's moratorium (Busch et al., 2015) or law enforcement and blacklisting in Brazil (Tacconi et al., 2019; Harding et al., 2021). A main difficulty of disincentive policies is that they tend to have lower political approval and imply higher political costs.

Finally, *incentive-based policies* include PES, integrated conservation and development initiatives (ICDP), and certifications schemes such as the Forest Stewardship Council (FSC). In the last two decades incentive-based projects have mushroomed across the world, PES as an example (Salzman et al., 2018). They have either taken place in a decentralised manner, with regional or international NGOs implementing incentive-based projects at subnational levels (Sills et al., 2014; Simonet and Seyller, 2015) or alternatively, as centralized and government lead initiatives such as PES in Mexico (Muñoz-Piña et al., 2008), Ecuador (de Koning et al., 2011), Perú (Giudice et al., 2019), China (Liu and Lan, 2015) or Costa Rica (Robalino et al., 2021). In general, such policies tend to have small positive or insignificant effects on both conservation and livelihood and welfare outcomes (Pirard et al., 2019).

Evidently, the impacts of each policy depend on context (Pfaff and Robalino, 2012). Two main reasons that affect policy performance are first, the economic pressure from deforestation differs across regions. Thus, the same policies might have lower effectiveness in areas or periods with higher opportunities costs of conservation (Chervier and Costedoat, 2017; Harding et al., 2021). Second, policies are not implemented in a "vacuum" or contexts that perfectly fit their theoretical application. Most often, policies are implemented simultaneously to other development and economic policies in which deforestation agents make decisions. Conservation policies also face changing economic conditions and institutional settings, which affect their transaction and implementation costs.

Paper I, II and III focus on an incentive-based policy that promises to reduce some of the transaction costs of conservation: collective incentive-based agreements. Collective PES – and collective agreements in general – are agreements that provide the payment or compensation to a group instead of an individual, based on the aggregated performance of individuals. They are a promising – and sometimes necessary policy, such as when land tenure is communal – that has caught more scholarly attention in recent decades (Hayes et al., 2019; Kotchen and Segerson, 2019). Compared to individual based PES, collective agreements offer three main advantages (Engel, 2016): (i) they are more suited to communal tenure arrangements or collectively managed resource systems[2], (ii) they can reduce transaction costs between 'users' of ecosystem services and the 'providers',

---

[1] Additionality is the capacity of the intervention to reduce deforestation compared to what would have happened without the policy.
[2] Up to 62% of land is communally managed or owned (RRI, 2015)

and (iii) they can improve ecological effectiveness and provision of ecosystem services via the conservation of larger areas. To examine the effectiveness of collective PES, the thesis draws heavily from behavioural economics and collective action theory.

1.2.3 Collective PES as a social dilemma

Collective PES can be framed as a collective action problem under two conditions: (i) when strict conditionality[3] is applied to the collective conservation outcome, and (ii) when it is impossible to exclude from the collective benefits the resource users who do not engage in conservation activities. In such cases, the decision to conserve the forest takes the form of a social dilemma. A linear social dilemma with extraction can be formalized with the following equation[4]:

$$\pi_i = x_i + \alpha\left(S - x_i - \sum x_{-i}\right)$$

π_i is the individual payoff, x_i represents an individual's extraction from the common-pool resource, x_{-i} represents the extraction of the other resource users. The common-pool resource yields a marginal return of α. The letter S represents the total stock of the common-pool resource, in this case, the forest under the collective PES. Let the letter N represent the total number of resource users.

To create a collective action problem, two conditions must hold: (i) the individual marginal returns from conserving the common-pool resource must be lower than the returns from extraction ($\alpha < 1$) and (ii) the marginal collective benefits of conservation must be higher than the marginal returns from not conserving ($N * \alpha > 1$). Participants of this social dilemma 'cooperate' when they decide to conserve the common-pool resource, thus maximizing collective earnings. In contrast, participants 'free-ride' when they decide to extract form the common-pool resource to maximize individual earnings at the expense of the collective (i.e., others') benefits.

The literature on collective action has highlighted key elements that affect cooperation rates in common pool resources. These can be broadly classified in four categories (Agrawal, 2001):

i) *Resource characteristics,* such as well-defined boundaries, mobility of the natural resource, and size of the resource.

---

[3] 'Conditionality' means that the payment is dependent on the conservation outcome. In many PES programs, conditionality is rather weak: payments are received regardless on conservation performance (Wunder et al. 2018).

[4] A social dilemma can be presented as either a Voluntary Contribution Mechanism (VCM) or a Common pool resource (CPR) game. The distinction relies in the decision task. In VCM situations, the participants decide on a contribution to a public good, and thus create a positive externality on other group members. In CPR games, the participant decides on extracting from the public good, thus creating a negative externality on other group members.

ii) *Group characteristics*, such as number of resource users, shared norms, past experiences of social capital, good leadership, interdependence among group members, and heterogeneity amongst resource users.

iii) *Institutional arrangements*, such as the possibility to sanction, monitor, and the relative ease of enforcement of the rules.

iv) *External environment*, such as having a low-cost exclusion technology, or supportive external sanctioning institutions.

Paper I and Paper II examine whether some of the elements of collective action theory outlined above behaved consistently and systematically across three country sites in Brazil, Indonesia and Peru. They explore how (ii) group characteristics, (iii) institutional arrangements and (iv) external environment affect cooperation in collective PES. More specifically, Paper I focuses on how inequality, or heterogeneity in endowments, affects the performance of PES when implemented under three alternative sanctioning strategies to mitigate the free-rider problem.

The three sanctioning strategies investigated are: (i) 'Public monitoring', which is a social, non-monetary sanction involving the public display of individual cooperation rates, (ii) 'Community enforcement', where community members can sanction their peers, and (iii) 'Government enforcement', in which an external enforcer intervenes to increase cooperation. Adequate monitoring, as well as the possibility to sanction peers, are two elements of the eight design principles for successful management of the commons (Ostrom, 1993). Hence, both 'Public monitoring' and 'Community enforcement' represent an 'Ostromian' approach to reduce free riding, while the last treatment has a 'Hobbesian' approach to enforcement, similar to command and control. Framed field experiments offer the opportunity to conduct a comparative study of these policies across countries. Only Vollan et al. (2019) has compared peer-to-peer and external sanctions using a framed field experiment in Namibia, and there are no evaluations comparing the effect of increased monitoring from the one of introducing sanctions.

The motivation to focus on how inequality in endowments affects cooperation under different institutional arrangements stems from the observation that inequality has been a relatively unexplored empirical question in the management of collective goods. While the basic theory regarding the impacts of inequality on common pool resources developed three decades ago, with the seminal work of Baland and Platteau (1999), the empirical evidence (particularly experimental evidence) exploring the issue has lacked behind. The impact of inequality on the commons is overall ambiguous, but in regulated settings inequality is associated to lower cooperation rates (Baland and Platteau, 1999). Exposure to inequality also increases risk taking attitudes and frequency of peer punishment (Kingsley, 2016; Payne et al., 2017). Thus, the effectiveness of enforcement mechanisms could be expected to be lower under conditions of inequality.

Paper I compares the performance of each sanctioning strategy in terms of effectiveness, efficiency and equity (3E) outcomes. Effectiveness refers to the degree to which deforestation is reduced. Efficiency is the degree to which the monitoring and enforcement increases the income of local forest users. Equity refers to the distribution of the earnings and the fairness perceptions. Empirically, the comparison is relevant to identify if one type of enforcement can substitute for the other and possible trade-offs amongst the 3Es. Theoretically, it is interesting to evaluate whether the performance of each of the free-riding mitigation strategies is conditioned on the exposure to inequality.

Paper II takes a more detailed look into the dynamics of internal, peer-to-peer punishment decisions. Participants who decide to punish their free-riding peers create a second-order collective good: the punishment might reduce future deforestation and thus increase future collective benefits. Motivations to punish free riding include altruism, fairness and equity concerns (Fehr and Gächter, 2002; Fehr and Fischbacher, 2004). Punishment can be also motivated by retaliation, spite and the wish of revenge (Fehr and Gachter, 2000; Herrmann et al., 2008) and thus not be targeted to the largest free riders.

The first research question of Paper II is thus to examine the patterns of punishment behaviour, trying to disentangle between punishments motivated by retaliation and revenge from those that have the Ostromian motivation of punishing free riders. It identifies behavioural typologies depending on the first order-cooperation (i.e., the contribution to the social dilemma), and second-order cooperation (i.e., whether resource users engage in punishment decisions). Identifying and understanding these broad behavioural typologies is important because they are correlated to forest conservation outcomes (Rustagi et al., 2010).

The second question addressed in Paper II concerns punishment effectiveness. Does the introduction of punishment change the marginal incentives to cooperate, and does it vary by site? To answer this question, the paper examines (i) the incentives to deviate from the social norm, defined as the generalized pattern of behaviour (i.e., the group average) and (ii) whether the deforestation levels change when there is possibility to punish.

1.2.4 Collective PES on the ground

Paper III "Land use and livelihood impacts of two collective incentive-based conservation interventions in Ucayali, Peru" investigates the impacts of collective agreements as they take place on the ground. The paper focuses on two incentive-based initiatives: an NGO-lead REDD+ project, and a government-funded collective PES, the National Forest Conservation Program (NFCP). Both initiatives are implemented in a manner far different from the ideal PES presented and addressed in Paper I and II. As many PES labelled initiatives, they have characteristics that make them more similar to ICDP projects than to the 'ideal' PES scheme (Wunder, 2015), including low enforcement, no conditionality, and payments delivered as in-kind assets for investment in alternative activities (instead of compensating for the opportunity cost of conservation).

The chapter presents a theory of change (ToC) for both interventions, inspired by the long-standing literature on microeconomic agricultural household models (Singh et al., 1986; de Janvry et al., 1991; Angelsen, 1999; Muller and Albers, 2004) and the livelihoods framework (Scoones, 2009). Such theories allow to identify key moderators that can affect both program participation and impact. An inflow of resources from the PES-ICDP projects can relax capital constraints that can in turn change the relative profitability of alternative sources of income, making the overall outcome for forest use and deforestation uncertain. The ToC presented in the paper thus remains ambivalent about expected program impacts on forest use.

The two programs are implemented in the same site, which provides an interesting opportunity to compare a “user financed PES” and a “government financed PES” and examine how design and implementation features affect both livelihood and wellbeing outcomes. Two key questions relevant for improving the understanding of collective PES are examined: (i) which households are most likely to actively participate in the program’s activities? and (ii) what are the impacts of the interventions on both participant and non-participant households?

A novelty of Paper III is that it focuses on anticipation effects (Malani and Reif, 2015), that is, how expectations about future program implementation affect current land uses and livelihoods. While theoretical and empirical analyses have pointed out that expectations at the start or during conservation projects can mobilize actors and resources (Harstad, 2016; Massarella et al., 2018), there are no impact evaluations testing for these effects. Expectations of future income can in fact influence the investment decisions of households (Aggarwal and Brockington, 2020), and a postponed or delayed payments can increase environmental degradation (Harstad, 2016).

1.2.5 Economic growth theory meets the forest transition

The final paper, Paper IV, moves beyond the local scale to examine the old question of drivers of deforestation at the national scale. The shift is justified by the fact that some of the variation in policy impacts observed at the local level are attributed to factors happening at a bigger scale, such as global commodity markets (Harding et al., 2021), international trade (Pendrill et al., 2019), or international agreements (Kerr and Policy, 2013). Indeed, one cannot see the forest for the trees: the impact of local interventions such as collective PES are shaped by broader trends, external to the local communities and actors. Paper IV thus aims to examine the broad trends driving national deforestation.

A theory traditionally used to explain changes in forest cover over time is the forest transition theory (Mather, 1992). According to the forest transition, national forest cover follows a U-shaped pattern: countries experience high deforestation rates at early stages of development, that decrease over time until forest cover eventually recuperates and increases. Rather than a theory itself, the forest transition is an empirical regularity. As such, it does not explicitly describe and model the assumptions or underlying causes of forest cover change. Nevertheless, three main pathways have been put forward to explain the forest transition, each suggesting different causal mechanisms to explain forest cover dynamics over time (Rudel et al., 2005; Lambin and Meyfroidt, 2010; Angelsen and Rudel, 2013).

The first is the 'forest scarcity path', where a decrease in the supply of forest products and ecosystem services over time leads to scarcities that raise the value of forests and incentivize forest recovery. The second is the 'economic development path', where forests recover as the result of higher off-farm employment opportunities, raising labour costs and thus decreasing incentives to deforest. The latter is the industrialization and modernization path, and it is akin to the environmental Kuznets curve: environmental degradation is reduced with economic growth. A third main type of pathway is the 'policy pathway', in which forest recovery is not driven by market and economic forces, but by changes in government policies and measures.

Old and recent studies have tried to identify the forest transition pathways and drivers of deforestation at the national scale, relying on broad indicators such as GDP and population density (Kaimowitz and Angelsen, 1998; Wolfersberger et al., 2015; Leblois et al., 2017). However, the principal weakness of many studies is that they lack an explicit theoretical framework, and thus mix 'direct' drivers of deforestation with its 'indirect' drivers. Paper IV introduces a more explicit theoretical framework to explain such changes.

Economists have for more than half a century explained output growth by either the factor accumulation or the growth of total factor productivity (TFP). In the basic Solow model, the aggregate production depends on two main factors: labour (L) and capital (K). In dual economy growth models, the economy is comprised of two main sectors, a 'modern' and a 'traditional' one, that are related by a common factor of production, labour. A central characteristic of dual economies is that each sector has different marginal returns to each factor of production (de Janvry and Sadoulet, 2016).

Paper IV expands the dual economy models to a three-sector economy that involves the forest sector as a third sector. Deforestation represents factor "deepening", in which previously underappreciated forest land is converted into higher value agricultural land (Hartwick et al., 2001). The forest and the agricultural sector share the same labour, and as in the dual economy model, it is assumed to be perfectly mobile across sector. If growth is a process in which the returns to forest are increasing relative to the agricultural returns, then it can be expected that deforestation rates will decrease in the process of growth.

The paper examines the extent to which dual economy growth theories and structural change in the economy are relevant to explain national deforestation rates and separate direct drivers from the underlying ones. While the link between structural change and deforestation has been either explicitly or implicitly made in previous studies (e.g., Foster and Rosenweig, 2003; Barbier and Bugas, 2014), it has not been tested empirically at a cross-country scale.

1.3 Methods and study sites

The following section presents the definition of causality adopted in economics and used in this thesis (section 1.3.1). The subsequent sections (section 1.3.2 to 1.3.4) present a description of the empirical strategies and study sites used for each paper of the thesis.

1.3.1 A brief definition of causality

Economic science has largely adopted the definition of cause derived from the counterfactual model of potential outcomes (Morgan and Winship, 2015). This definition of causation focuses on the 'difference-making' properties of causes. An effect is defined with reference to a counterfactual scenario, in which the outcome variable Y is compared in situations in which the cause D is present and situations in which it is absent. Each individual in a population of interest can have two alternative states: one *without* the cause, and one *with* the cause D present. The alternative states, or 'potential outcomes', can be labelled Y_0, in which the cause is not present (D=0), and Y_1, which is the potential outcome in which the cause D is present (D=1).

The Average Treatment Effect (ATE) for the observations that received the treatment (i.e., the cause) is calculated as $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$, and for the observations that did not receive the treatment the effect is: $E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0]$. The fundamental empirical problem of this view of causality is that in the real world, for any unit of analysis, one can only observe one of the potential alternative states: where either the cause D is absent or in the case where D is present. That is, only $E[Y_{1i}|D_i = 1]$ or $E[Y_{0i}|D_i = 0]$ are observable, and it is impossible to observe $E[Y_{0i}|D_i = 1]$ or $E[Y_{1i}|D_i = 0]$. A naïve calculation of the average effect from observed data is:

$$ATE = E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

The naïve calculation can be misleading because each state (D=1 or D=0) is characterized by a distinct set of conditions which potentially affect the outcome of interest Y, or the probability of receiving the treatment D. That is, there might be pre-existing differences between the observations that received the treatment and those that did not. When the probability of receiving the treatment differs or there are covariates that systematically differ between treated and untreated units, there is 'selection bias' and the ATE cannot be derived directly form observational studies, because the potential outcomes between the treated and untreated observations is different. That is, $E[Y_{0i}|D_i = 1] \neq E[Y_{0i}|D_i = 0]$.

It is only when $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$ that one avoids selection bias in observational studies. A basic methodological challenge to establish causation is therefore to identify a proper comparison group (counterfactual) from observed individuals in which the cause of interest D is not observed.

The ideal method to identify causality are experiments (Athey and Imbens, 2017). Experiments are considered the gold standard because they have the property of manipulation. In experiments, by randomly assigning (i.e., manipulating) a treatment, the experiment can control for unobservable bias, by assuming that the confounding covariates are, on expectation, the same across treated and non-treated individuals. Paper I and II use the experimental method, in particular Framed Field Experiments (Harrison and List, 2004).

When manipulation is not possible, one must construct the counterfactual scenario using observational data with a 'quasi-experimental' approach. Some of the commonly used methods to construct counterfactuals from observational data are matching, propensity score, difference and difference, instrumental variables, and regression discontinuity (Athey and Imbens, 2017). Each method can control for either observed or unobserved differences between treated and untreated units.

Paper III uses propensity score matching with difference in difference to control for potential sources of bias. Paper IV in turn, relies on panel data which allows to control unobserved characteristics that do not vary in time or by unit of observation. It is important to note that not all the relationships contained in the thesis are causally identified. To make such distinction, and similar to Meyfroidt (2015), throughout the Papers the word 'treatment' and 'impact' are used to refer to 'causes', that is the relationships where, to the best of the methods and data, it has been possible to establish a causal relationship. The words 'determinant' or 'driver' are used whenever relations are statistically established, but where causality is not necessarily asserted (but not excluded either). This distinction in terminology is not often made in the literature but is useful for this thesis.

1.3.2 The Framed Field Experiment

1.3.2.1 An overview of the experimental design

The Framed Field Experiment (FFE) experiment used for Paper I and Paper II was designed similar to a Common Pool Resource game, framed as the decision to deforest an area under a collective PES. The game comprised 6 participants and had 4 stages with 6 rounds each (24 rounds in total). The first stage was the baseline stage, in which the collective action dilemma was introduced as a collective PES payment. Only average group deforestation was shared with the group.

In stages 2, 3 and 4, Public monitoring, Community enforcement (peer punishment), and Government enforcement (external punishment) were sequentially introduced. This means that the treatments followed a within subject design (two or more treatments for each group), which allowed to keep subject characteristics constant across experimental treatments. A within design is more appealing if high behavioural variability is expected across individuals or groups, relative to the variability caused by sequencing (Holt, 2007). Public monitoring was introduced in the second stage and maintained throughout stages 3 and 4 (i.e., from rounds 7 to 24, all participants could observe others' individual deforestation). The Government and the Community enforcement were played in stage 3 or 4, acknowledging that there could be strong sequencing effects between these two treatments.

Inequality in endowments was introduced as a between group design, by varying the maximum capacity to deforest of the participants. A between subjects design (one treatment per group) is better when there is less variability across individuals or groups, and when there are sequence effects that cause behaviour in one treatment to be influenced by what happened in an earlier treatment (Holt, 2007).

In half of the experimental sessions, three randomly selected participants had a high capacity to deforest up to eight plots, while the other three had a low deforestation capacity and could deforest only up to four plots. In the other half of the experimental session, participants had a medium deforestation capacity of six plots. The incentives for cooperation are the same for every participant, the only source of inequality is the 'deforestation capacity'. Any observed differences between the two treatments arises from the exposure to the randomly assigned heterogeneity.

The experiment had thus a 2x2 complete factorial design, in which the first factor is whether the group has Equal or Unequal deforestation capacity amongst participants. The second factor is whether the Government or Community enforcement is introduced first in the sequence of within-group treatments. This design resulted in overall four different types of experimental sessions, each experimental session type was conducted 30 times, 10 times in each country (Figure 1.1).

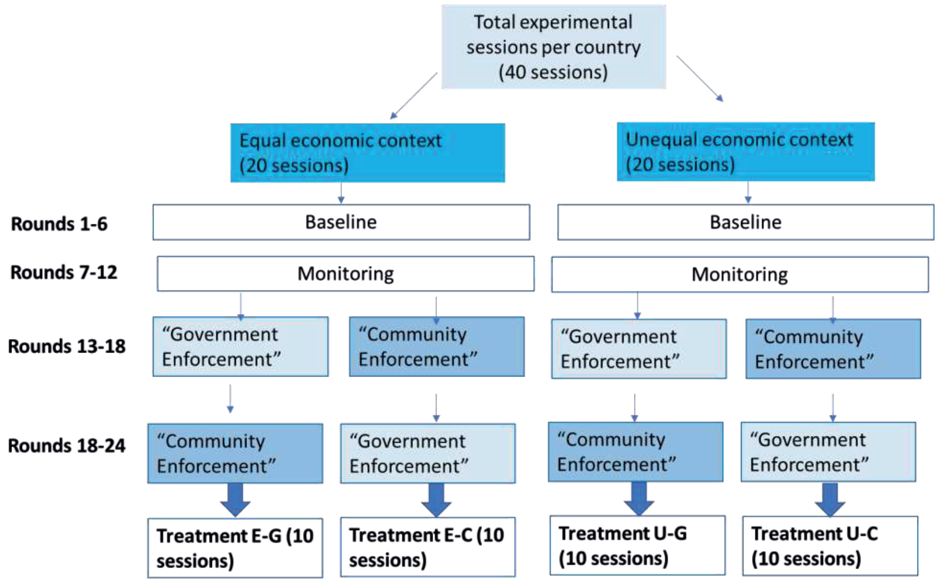


Figure 1.1. Overview of the experimental design.

There are limitations to the experimental design. First, there is no 'zero treatment' control group, in which neither Public Monitoring, Community and Government enforcement were played. The validity of the causal interpretation of our treatments thus relies on the unverifiable assumption that learning effects are small compared to the treatment effects, i.e., the first six baseline rounds represent a good counterfactual. Such experimental group was sacrificed in order maximize the number of treatment observations and focus on within group differences. By doing so, a rather tedious experimental session where participants played 24 rounds with exactly the same rules was also avoided. Even though one cannot completely control for learning effects throughout the game, three measures were taken to mitigate spillover effects: (i) the forest stock was reset to the initial level of 60 plots in every round, (ii) round controls were included in all regression analyses, and (iii) the identity of each player throughout the stages was changed (by introducing different letters for each player, in each stage).

Second, the situation framed in the experiment does not perfectly recreate the real-world PES. In the ‘real world’, collective PES are implemented in contexts where both internal and external enforcement are simultaneously at play to different degrees, and PES are imperfectly monitored and enforced (low conditionality). This choice of design and treatments stems from a specific conceptualization of what experiments can be useful for. Experiments can be used to (i) test theories, (ii) search for empirical regularities, and (iii) to generate policy advise[5] (Roth et al., 1988). Given that the experiment was to be implemented in three countries with different social, political and cultural contexts, there was more potential of using the experimental method as a tool to test theory and search for empirical regularities than as a tool for specific policy advise.

Thus, there was less concern for having perfect ‘treatment validity’ (Handberg and Angelsen, 2015), that is, the relationship between the experimental treatments and the actual policy, than to have interesting treatments in which to examine and identify regularities. In that sense, the design adheres more to the ‘scientific view’ of what experiments can be useful for, rather than to the ‘policy view’ (Camerer, 2011). While in the policy view, treatment validity is important, for the scientific view, understanding or identifying general principles is more important.

1.3.2.2 The three country sites in Brazil, Indonesia and Peru

Previous cross-country experiments repeatedly demonstrate that there are important variations across sites (Herrmann et al., 2008). Hence, to understand cross-country results from framed field experiments, proper identification and description of the social and economic context is important, particularly in the study of common pool resources (Anderies et al., 2011). The three sites selected for the study are in Pará (Brazil), Central Kalimantan (Indonesia) and Ucayali (Peru). The data was collected in autumn 2019, and the fieldwork was coordinated by different supervisors in each country.

The three sites have interesting characteristics that make them relevant for a comparison of the effects of a collective PES under different sanctioning institutions. Forests are owned communally in Peru, in Indonesia the land is owned by the state, while at the site

[5] These three potential uses are what Alvin E. Roth calls “speaking to theorists”, “searching for facts”, and “whispering in the ears of princes” (Roth 1988).

in Brazil land is owned individually by colonist farmers. In Peru and Indonesia, households' control, on average, an area of ~2.0 ha for subsistence and commercial agriculture, while in Brazil, households control, on average, an area of 44.8 ha of forest and 38.7 ha of agricultural land, mostly pastures.

In Brazil and Peru land tenure is in most cases considered secure, in the sense that collective and individual boundaries of properties are legally recognized. On the contrary, tenure is considered weak in Indonesia because village and households do not have legal recognition of the land they manage and forest access is based on local customary laws, which give individuals land claim when they have invested on that land (e.g., planting, clearing land) (Sills et al., 2014).

In each village, households who had previously participated in CIFOR's Global Comparative Study were prioritized to participate in the experiment. A total of 120 experimental sessions took place, 40 in each country. At least half of the villages in each country have received or are in the process of receiving monetary incentives for forest conservation, while also facing regulations and enforcement mechanisms.

1.3.3 Impact evaluation

1.3.3.1 Matching with differences in differences and the subjective measures

The quasi-experimental method utilized for Paper III relies on simulating randomness (exogeneity of treatment) by using propensity score matching methods with differences-in-differences. Without randomization researchers have to control for all potential variables that can affect treatment assignment or outcome variable, as there are potential confounders that can generate biases. To give reliable causal evidence an essential (unverifiable) assumption of matching methods is to satisfy the conditional independence assumption, which states that given a set of observable covariates X that are not affected by the treatment, the potential outcomes Y are independent of treatment assignment. This means that the treatment is as if random, conditional on the covariates:

$$(Y_{1i}, Y_{0i}) \perp D_i | X_i$$

The conditional independence assumption implies that we need to observe all the factors (confounders) that are correlated with both the outcome, Y, and treatment assignment, D. The second important identification assumption of matching strategies is the common support assumption, which states that there is a similar distribution of covariates between the control and intervention observations.

When treated and untreated units differ in unobservable characteristics associated to potential outcomes, there might be still subject to bias after matching. Thus, Paper III combines matching with the difference-in-difference to control for unobserved heterogeneity that does not vary over time. The method allows to identify effects by focusing on the differences between the control and treatment observations, before and after treatment implementation. The estimate of the average treatment effect on the treated (ATET) is thus:

$$ATET = E(Y_{1i}(t) - Y_{1i}(t-1)|D = 1) - E(Y_{0i}(t) - Y_{0i}(t-1)|D = 0)$$

$Y_{1i}(t)$ is the outcome variable after treatment, and $Y_{1i}(t-1)$ is the outcome variable before the treatment is implemented. The validity of difference-and-difference relies on the parallel trend assumption, which states that the trends in the treatment and control observations are the same, that is:

$$E(Y_{0i}(t) - Y_{0i}(t-1)|D = 0) = E(Y_{0i}(t) - Y_{0i}(t-1)|D = 1)$$

Given the collective nature of the incentive-based interventions considered in Paper III, and the observation that the participation in incentive-based interventions varies within villages, the paper focuses on evaluating the treatment effects on different pools of participants: it examines the Intention to Treat (ITT), the average treatment effect on the treated (ATT), and the average treatment effect on the non-participants (which we label 'ATNP').

The programs are evaluated using the quasi-experimental method outlined above, as well as the personal perceptions of respondents, referred to as 'self-reflexive evaluations' (Schreckenberg et al., 2010). The inclusion of subjective measures is justified on two fronts. First, objective measures such as income and reported land use are not enough to understand impacts: arguably, perceptions of the participants are just as important or even more so for both current and future land use and program participation. Second, subjective evaluations bring out procedural aspects such as participation and transparency during project implementation, which are an important component of REDD+ safeguards (Duchelle et al., 2017). By using both the objective and subjective measures, the paper provides a more comprehensive approach to program evaluation.

1.3.3.2 The Ucayali site in the Peruvian Amazon

The study site comprises eight villages belonging to the Shipibo-Konibo indigenous group, all located in the Ucayali district of the Peruvian Amazon. Peru is the second country with the largest cover of the Amazon rainforest, covering approximately 12%. The Ucayali district is the second district with the highest terrestrial carbon (Asner et al., 2014; Csillik et al., 2019) and the highest deforestation rates in Peru (MINAM, 2020).

The drivers of deforestation in the area are mostly related to smallholder agricultural expansion, with large scale oil palm playing an increasing role (Bennett et al., 2018). The ecological conditions in which these villages are located, as well as their distance to urban markets, constrains households' primary livelihood strategies to agricultural, fishing, and forest extraction activities (Coomes et al., 2010; Rodriguez-Ward and Paredes, 2014; Porro et al., 2015; Begazo Curie et al., 2021).

Participation in the labour market is seasonal and sporadic. Three villages are in seasonally flooded forests with limited potential for agricultural expansion, while five villages are in non-floodable forests and thus more suited for agricultural activities. Farming, fishing, and agricultural activities occur mostly during the dry season, which runs from April to September.

1.3.4 Fixed effects

Paper IV exploits the benefits of panel data to analyse macroeconomic indicators. Panel data consists of repeated observations over time of the same units of analysis or entities. It is a powerful yet relatively simple method which allows to control for many omitted variables. The combined unit and time-fixed effects models allow to control and omit biases from unobserved variables that are constant over time (similar to the difference-in-difference model), and from unobserved variables that are constant across units of analysis. The basic model specification is as follows (Stock and Watson, 2015):

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \gamma_t + u_{it}$$

Where α_i is the unit fixed effect and γ_t is the time fixed effect, and u_{it} are standard errors for each unit. While two-way fixed effects models control for the constant heterogeneity across time or units, a potential threat to the identification strategy is to control for variables that vary both across space and time. When applied to the analysis of deforestation, this means that one must include as control variables that can vary both over time and across units of analysis.

The first step in the empirical approach of Paper IV is to estimate the marginal products of each factor of production in both the agricultural and modern sector. The standard way to estimate macroeconomic returns to capital is with a production function (translog or Cobb–Douglas, generally) and to estimate the main elasticities of each factor of production using OLS regression. The second step in the empirical approach of the paper is to examine how the relative changes in the marginal returns of each factor of production in each sector (i.e., the immediate drivers of deforestation) affect deforestation rates. The final step is to examine the determinants of the changes in the relative returns of the factors of production, these being the 'underlying' drivers of deforestation.

The main difficulty of using fixed effects models as a clear identification strategy is that they do not control for the bidirectional nature of the causal relationship (i.e., simultaneity bias). In the case of deforestation, a positive coefficient with economic growth can indicate that deforestation causes growth, just as well that growth causes deforestation. Potential solutions include dynamic panel data analysis (Arellano and Bond, 1991; Roodman, 2009), or including lagged independent variables, but it does not completely resolve for the issue (Leszczensky and Wolbring, 2019). Difference GMM can be useful when exogeneity is not met, as it constructs instrumental variables based on difference in time lags of independent variables. None of these approaches are used in Paper IV, thus throughout the chapter there is a conservative interpretation of the results and the language used to name the independent variables remains "drivers" or "determinants" instead of causes.

1.4 Data

The following sections describe and discuss the data collection process. Three types of data were used in the thesis. The first is experimental data (Paper I and Paper II), the second is survey data (Paper III), and the third is secondary data (Paper IV).

1.4.1 Experiments and experimental data

Experimental methods have become a central method in economics (Harrison and List, 2004) and in the study of common-pool resources ever since the first experiments with Ostrom (Ostrom et al., 1992; Ostrom, 2006). In experiments individuals are extracted from their 'natural' environment and transported to a controlled setting which is easier for the researcher to understand and trust. As a result, experimental data is relatively straightforward to collect and analyse. From the simplified cases one obtains an idea of what can occur in more complex situations, as well as identifies general principles and patterns of behaviour. However, by reducing and abstracting a problem from reality, it is harder to capture the precise magnitudes of the effects that could be observed in natural environments (Kessler and Vesterlund, 2015).

An important question of scientific studies in general, and experimental studies in particular, is the extent of external validity, defined as the degree to which the results of an experiment can be transported and generalized to other (natural or not) environment or population (Camerer, 2011). The question of external validity is speculative by nature: it is difficult to test, and it crucially depends on what aspects of the experimental results want to be applied to another setting[6]. Hence, the evidence about external validity of experimental results is mixed. Some studies find a correlation to real life situations (Rustagi et al., 2010), while others find none (Voors et al., 2012; Galizzi and Navarro-Martínez, 2019).

Nevertheless, it is possible to outline factors that are likely to affect the external validity of the results. Levitt and List (2007) point out five important factors that affect the generalizability of experiments to other contexts and populations: (i) the presence of moral and ethical considerations, (ii) the nature and extent of scrutiny of one's actions by others, (iii) the context in which the decision is embedded, (iv) self-selection of the individuals making the decisions, and (v) the stakes of the game.

Framed field experiments as conducted for this thesis, compared to lab experiments, mitigate a couple of the issues pointed out by Levitt and List (2007). First, by considering actual forest users instead of self-selected university students – the WEIRD subject pool (Henrich et al., 2010b) –, it increases external validity as participants bring their own ethical considerations and relationships into the game (Cardenas and Ostrom, 2004). In addition, by framing the game according to the relevant decision context or 'domain' in

[6] As succinctly explained by Camerer (2015): "It is certainly true that *some* economics experiments have features that make them less likely to generalize to *some* natural occurring settings. However, the exact same thing is true of *some* field settings that will not generalize well to *some* field settings" (p.251).

which the decision needs to be taken (in our case, a forest management context), it allows to better identify and interpret the actual behavioural choices of participants. This allows to increase the external validity and generalizability of the results to the actual context and increase experimental data's usefulness for policy advise (Harrison and List, 2004; Anderies et al., 2011; Finkbeiner et al., 2018).

There are at least four remaining important aspects that still threaten the external validity of the framed field experiments, particularly for policy purposes. First, the nature of the stakes. Lower stakes increase the chance of having random answers (just as hypothetical payoffs do), and they can increase the willingness to cooperate in a social dilemma. Higher stakes have been shown to decrease trust levels (Parco et al., 2002), or contributions in dictator games (Carpenter et al., 2005). In the FFE used for Paper I and II, this is best evidenced by the 'random' nature of some of punishment decisions, where participants declared that they punished just because "they could". Unfortunately, there is no measure that can satisfactorily mitigate this problem without significantly increasing the costs of the experiment.

The fact that one must operate with relatively small stakes is a reason why recreating a non-cooperative environment, without communication, was preferred for the experimental design used in Paper I and II. Presumably, if the experiment captures individual motivations in a low stake, non-cooperative environment, one could venture to say more about situations where stakes are much higher, and thus coordination and cooperation harder.

A second aspect that threatens external validity is the nature of the scrutiny. The scrutiny of one's own actions are different in natural as compared to experimental settings for two reasons. First, experiments are 'obstrusive' by nature (Camerer, 2011), meaning that participants know that their actions are being observed and might be judged by the experimenter. Thus, they might choose to behave strategically towards the experimenter by cooperating more. Second, there is scrutiny originating from their peers. One cannot assume that the controlled setting of the experiment, where participants only interacted with five other peers under conditions of anonymity, adequately represents all the factors that participants would consider when making the decisions in their real life. In real life, participants might be subject to more (or less) scrutiny by their peers than what they are subject in the experiments.

A third aspect for the external validity is, naturally, the time horizon that is presented in the experiment compared to that of the real word. During the experiment, participants make the decision in a short-run (two hours), 'hot' context (Levitt and List, 2007), while actual participation in PES program involves 3-5 year commitment, and participants have more time to analyse and communicate their decisions and opinions (i.e., 'cold' decision making).

A final consideration, relevant for both internal and external validity, is the extent to which participants fully understand the procedures and implications of their decisions.

The manner in which the problem is presented, and the environment created shapes the cognitive abilities of respondents. To increase the understanding of the social dilemma, it is common practice in framed field experiments to provide a visual support representing the forest managed under the collective PES, give examples, answer questions, and indicate the payoff loss/gain when an individual decides to deforest, as well as their corresponding payoff tables.

1.4.2 Household-level survey data

Collecting socioeconomic data of rural households is important and necessary to understand forests' ecosystem services, the relationship between poverty and forests, the costs of forest conservation, and the impact of policies such as ICDPs and PES. Historically, the breath of ecosystem services and values from nature has been largely underestimated: the contribution of natural capital to rural households' income has been the 'hidden harvest' (Scoones et al., 1992). There has been little understanding on the relationship between poverty and environment, and how do forests contribute to environmental and rural livelihoods.

One of the first studies measuring the income benefits of environmental resources was by Cavendish (2000) in Zimbabwe, which showed that the environmental resources could make up to 40% of poor rural households income. Years after emerged the Poverty Environment Network, a global study to measure rural livelihoods (Angelsen et al., 2012). The PEN study focused on the quantification of environmental income, as well as in the understanding of the links between poverty and forest reliance in developing countries. CIFOR's GCS REDD+ study is an inheritance from those previous efforts, and has made significant contributions to the identification of the impact of conservation policies in the REDD+ context (e.g., Bos et al., 2017; Sills et al., 2017; Simonet et al., 2018), as well as identifying smallholders' opportunity costs of conservation (Ickowitz et al., 2017). The GCS REDD+ household survey follows income measures[7] and consists of an approximately two-hour long survey to household heads (Sunderlin et al., 2016).

In September-November 2018, socioeconomic and demographic data from 247 households of the Shipibo-Konibo indigenous group in Ucayali, Peru, was collected with seven enumerators. Indigenous populations in Peru such as the Shipibo-Konibo have shifted from being hunters and gatherers, to swidden cultivation, to increasingly having more sedentary agriculture at the forest frontier. The Shipibo-Konibo population of rural households is particularly interesting (and perhaps difficult) population to survey because even though the share of forest income is being reduced over time, households still have a relatively high degree of environmental and forest income compared to other GCS REDD+ sites (Sills et al., 2014).

---

[7] This approach contrasts with the expenditure approach to measuring rural household income. Yearly expenditures are more stable, but the income approach is necessary to establish a link between poverty and forest use.

The data collected can be judged by accuracy and precision. Accuracy refers to the capacity of the data to capture the 'true' values, while precision refers to the distribution and dispersion of the data collected (Jagger et al., 2012). Low accuracy implies bias, while precision is tied to the variation in the data. When collecting survey data, there are many sources of errors that reduce accuracy and precision, but also simple ways to mitigate them (Lund et al., 2011).

First, errors arising from the enumerators or the imprecise nature of the questionnaire. To mitigate this source of errors there was a four-day detailed training with local enumerators to carefully revise the questionnaire, and during the data collection period they worked in pairs as much as possible. Enumerators that had participated from previous phases of the GCS REDD+ project (in 2012 and 2014) were hired, and selected to interview the same households, so that the traceability and trust of the respondents towards the enumerators would increase. Thus, enumerators had previous knowledge about the research objectives and familiarity with the methods and questionnaire.

A second source of concern is the extent to which households engaged in strategic answers. As enumerators, one can try to minimize the errors by recalling the independent nature of the research at the start of each survey, and the that it is for research purposes only, and in their best interest to answer honestly to better understand the impact of the interventions. Strategic bias is a greater source of concern for variables such as reported deforestation and forest use, than for example, durable assets and household size. The cooperative behaviour of the respondent is assumed, but cannot be guaranteed (e.g., that they mention all the monkeys and boars they killed as a source of bushmeat). The setting in which the survey took place - in most cases at the households' home -, as well as the fact that the enumerator team stayed for 4-5 days in each village, allowed to triangulate the data a lot of the data (e.g., confirm the durable assets or cross check information that seemed to be an outlier).

A third and final aspect that could decrease the precision of the measurements are the cognitive abilities and bounded recalling capacities of households. In some cases, it was hard for households to recollect different sources of income, with some households even forgetting the number of children! This source of error can especially happen when there are long recall periods such as one year, as used in the survey. To help households recall their income, the questionnaire decomposes the income stream by different sources (e.g., income from fishing, from logging, etc.), and different components (e.g., costs, expenditures, seasonality).

The above-mentioned sources of errors reduce accuracy and precision, potentially introducing bias and increasing standard errors. If strategic bias is reduced and minimized, the measurements errors in dependent variables produce unbiased but inefficient estimates of the parameters, which might reduce the statistical significance of the results. Thus, measurements errors are more likely to accept null hypothesis, and increase the likelihood of obtaining a false negative, which might influence the results of Paper III.

While large standard errors in the data can be partly attributed to the elicitation method, they are also the consequence of the reality that the data aims to describe. Rural households have unstable sources of revenues, with important year to year oscillations, and face windfall income opportunities as well as dependence on climate conditions. Acknowledging these data limitations, and given the costs and time constraints, the survey conducted is nevertheless one of the best possible sources of information to understand livelihood changes, household level policy impact, and the poverty-forest relationship.

1.4.3 Secondary data

When dealing with secondary data, an appropriate contextualization and understanding of the data generation process is important to have correct interpretation as well as estimate possible biases during data analysis. While the quality of macroeconomic data has significantly increased over time, it is still subject to important biases and measurements errors (Jerven, 2013). This section briefly presents how the macroeconomic indicators used in Paper IV of the thesis are produced.

1.4.3.1 Capital stock indicators

Two indicators of capital are used, one for the modern sector and one for the traditional sector. The capital of the modern sector is calculated using the perpetual inventory method (Caselli, 2005), which consists of estimating the capital stock at time t based on the previous depreciated capital and current investment[8]. The capital stock is thus based on all previous investments leading up to that year. An important challenge is to estimate the capital stock in the first year of the data. PWT 9.0 assumes that when a country's data is first observed, its nominal capital-output ratio is 2.6 (Feenstra et al., 2015).

To measure capital stock of the traditional sector, the FAO's Agricultural Capital Stock was used. Similar to the Penn World Table, it is calculated using the perpetual inventory method. It is an analytical data base, in the sense that it integrates data from national accounts collected by the UN's statistics division or OECD (Vander Donckt et al., 2021), and following the definition of assets of the System of National Accounts (SNA)[9]. An important element for the purposes of the analysis in Paper IV is that in the SNA definition for capital stock non-produced assets are excluded from the fixed assets. Thus, non-produced assets that occur in nature, such as natural resources, subsoil assets, or non-cultivated biological resources (i.e., forests), are not included in the capital accounting. Because a full set of official data is not available for all countries in all years, when data is deficient, the Statistics Division at the FAO relies on imputation methods to generate a complete time series.

---

[8] The specification is: $K_t = I_t + (1 - \delta)K_{t-1}$

[9] System of National Accounts (SNA) is the internationally agreed standard set of recommendations on how to compile measures of economic activity, of the United Nations Statistics Division.

1.4.3.2 Labour indicators

For indicators of labour, sectoral indicators of employment as reported in the World Development Indicators (WDI) are used. These indicators obtain the information from the International Labour Organization (ILO). Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job. Two main limitations of the data that are not considered in Paper IV, is that the data on employment does not consider the quality of the labour force by including, for instance, educational indicators. Thus, it is an incomplete indicator of the human capital component of a national economy. A second limitation is that the size of the informal sector in developing economics is high, and that self-employment is not reported in the ILO database. As a result, the employment share can be underestimated, potentially leading to overestimation of labour productivities.

1.4.3.3 Land use and deforestation indicators

The data on forest and agricultural land come from the Food and Agricultural Organization (FAO). The FAO collects annual data of land use through country questionnaires. In case of missing information, gap filling, imputations are performed by FAO in line with the International Statistical Standard. Because the data comes from country questionnaires, the geographical comparability may be limited, due to differences between countries in methods, definitions, and coverage. Furthermore, the degree of accuracy is not assessed. The accuracy and validity of the data is dependent of the country's statistical capacity (Keenan et al., 2015). In the most extreme cases, FAO's statistical office must adjust land values so that the sum of the land categories is equal to the total country area.

The deforestation data that is used comes from Hansen et al. (2013), which is the most commonly used data set in current deforestation studies. Yet, the data set is deficient to examine the forest transition and deforestation rates for two reasons. First, it only describes forest cover loss, and not net change in forest cover at the country level. To the extent that the paper focuses only on explaining deforestation rates, this is not a source of major concern. More important is that the data considers as forest cover loss the clearing of tree crops, such as palm oil. Thus, the data does not provide an exact indication of agricultural expansion.

1.4.3.4 Governance indicators

The World Governance Indicators (WGI) are the most frequently used indicators of governance for cross-country comparisons. The indicators are built using multiple data sources which include surveys to measure perceptions of households and firms, businesses, non-governmental organizations, and public sector organizations. They are reported every 1 or 2 years, adjusted to have a zero mean and unit standard deviation in each period. Each indicator is composed of multiple weighted variables (Kaufmann, 2007).

Three aggregate indicators are used as controls in Paper IV: Rule of Law, Regulatory Quality, and Political Stability. These are commonly used indicators in the analysis of deforestation (Barbier and Tesfaw, 2015; Leblois et al., 2017). The indicators have been criticized for being too imprecise, with different sources of underlying data, or biased towards views of business elites (Kaufmann, 2007). Some of these criticisms are explained by the difficulty of measuring and defining governance itself. Although imperfect, they are the most comprehensive and reliable indicators for cross-country comparisons (Arndt and Oman, 2006).

1.5 Main findings, contributions, and limitations

1.5.1 Inequality and enforcement in a collective PES.

Paper I “Enforcement and inequality in collective PES: effectiveness, efficiency and equity implications” shows that introducing monitoring and enforcement allows to significantly increase the conservation benefits of collective PES, although the impacts vary across the three country sites. Public monitoring of individual decisions has limited effectiveness as compared to the introduction of monetary sanctions, and a significant effect is only observed in sites with a stronger history of collective action (Peru and Indonesia). Community enforcement increases effectiveness but can reduce the efficiency and equity of collective PES. Government enforcement provides the strongest and most robust results in terms of effectiveness and efficiency outcomes. Due to the random nature of Government enforcement, and the ‘overuse’ of peer punishment during the Community enforcement, none of the punishment strategies increases the distributional equity aspect of PES.

The experimental findings are discussed in light of the three potential contributions that economic experiments can generate (Roth et al., 1988; Roth, 2015): to test theories, to generate facts, and to generate policy advice. Regarding the first, to test theories, the research is a significant contribution to the relatively small experimental literature that deals with heterogenous agents. It shows how institutional performance can be affected by the existence of inequality in endowments, but perhaps not as strongly as one could expect, as it only had an impact in Indonesia. Results indicate that the effect of inequality in endowments is site dependent, serving as a warning about generalised statements about the impacts of inequality of deforestation and policy effectiveness.

A second important contribution of the experimental results relies in generating interesting facts that need further explanation. Why did the inequality treatment have a strong effect in Indonesia but was absent in the Latin American sites? Rather than individual level characteristics, it is claimed that these differences originate from contextual factors such as land tenure regimes and land tenure security. It is, however, an unsatisfactory explanation of the results as it cannot be tested directly with the current data. Future research should look at within country variation and test interactions of inequality with individual level characteristics such as social preferences or village level characteristics, such as market integration or forest dependence.

As policy advice, two important general messages that are to be taken from the results of Paper I are that unless collective PES is accompanied by the possibility to sanction, it is unlikely to be very effective in the Brazilian country site. This result is attributed to the individual land tenure, and less history of collective action in the site. On the other hand, the results serve as a warning and a ‘proof of concept’ of the potential dangers of collective benefits when land tenure is insecure: they could create conflicts and disagreement amongst community members, as the results of lower efficiency in Indonesia testify. This is consistent with basic theory of collective action: one of the eight design principles for successful management of the commons is to have clearly defined boundaries (Ostrom, 1993; Wilson et al., 2013).

Regarding the limitations, an interesting aspect could have been to include questions about fairness norms, as evidence shows that the origin of wealth differences affects fairness perceptions (Almås et al., 2010). Further, one is also left wondering if the inequality effect would have been stronger with a stronger treatment, such as with a real incentives task for example (Loft et al., 2020). Finally, examination of the impacts of other sources of inequality are an interesting avenue of research, for instance, by mixing inequality in endowments with inequality in returns from the public good, *à la* Hauser et al. (2019), or interacting inequality with different PES distributional rules (Vorlaufer et al., 2017).

1.5.2 Peer punishment in a collective PES regime.

A central finding of Paper II “Will peer punishment protect tropical forests? Multi-country evidence from framed field experiments” is that peer punishment can deliver on conservation outcomes and reduce deforestation in the context of collective PES. Similar to the findings in Paper I, the results show that the effectiveness of the enforcement to increase cooperation varies by site: it is strongest in Indonesia. Further, there is a consistent relationship between first order cooperation and second order cooperation: the participants who extract less from the CPR are also the ones who are more likely to punish as well. Mirroring that, there is a consistent relationship between free riding in the first social dilemma and giving antisocial punishment.

Results also show that the severity of the punishment depends on the size of the violation of the social norm. Participants who deviate more from the social norm of average group deforestation, are likely to receive more punishments. This is particularly true for the Indonesian site, and overall consistent with the graduated sanctions criteria of successful collective governance (Ostrom, 1993). While receiving punishment does effectively reduce deforestation, self-enforcement entails a risk of engaging in antisocial behaviour which – besides being costly in itself – has a negative effect on future cooperation.

Future examination of how patterns of antisocial and prosocial punishment evolve over time will help increase the understanding of peer punishment dynamics. In addition, an interesting extension would be to examine if the patterns of peer punishment found in this study remain if peer punishment is introduced with a coordinating device such as communication (Gangadharan et al., 2017), or voting on punishment (Pfattheicher et al., 2018; Nockur et al., 2021).

1.5.3 Collective PES on the ground

The results of the quasi-experimental evaluation in Paper III "Land use and livelihood impacts of two collective incentive-based conservation interventions in Ucayali, Peru" show that the National Forest Conservation program (NFCP) and the REDD+ projects are not significantly affecting income, and there is suggestive evidence that they can have positive effects on the amount of households' forest land used, but that this is not necessarily explained by changes in the expectations, as was originally hypothesized. A main contribution is thus to show that despite the slow and delayed implementation of the conservation projects, there are no negative behavioural changes or 'conservation hostages' (Harstad, 2016).

The study also brings forward the advantages of considering household self-reflexive evaluations, or perceptions, to both understand how an intervention might be having 'intangible' positive effects on wellbeing, such as increased monitoring of common forest, as well as identify how procedural practices during program implementation affect household's evaluation. This is important because perceptions can affect future and continued participation. The chapter shows that the quality in which information is transmitted to the communities *does* matter, as well as the transparency of the process. It also highlights the importance that participants attribute to increasing their forest monitoring capacity.

The chapter has two main limitations. First, the definition of program participation is rather weak. The early stage of implementation of the programs, in particular of the REDD+ program does not allow to clearly identify beneficiaries of the program, which explains why there is no significance of explanatory variables examining participation rates. Second, there is no way to test for the mechanism presented in the Theory of Change (ToC) of the chapter. More specific questions about the households' expectations about the program could have disentangled between the potentially opposite effects of expectations on land use and livelihood variables.

1.5.4 Re-examining the macroeconomic drivers of deforestation

Drawing from dual growth models, Paper IV "Re-examining the macroeconomic drivers of deforestation" develops a theoretical framework to analyse the drivers of deforestation and disentangle the immediate drivers form the underlying drivers. Results indicate that the constructed indicator of structural change, the relative returns of forest to agricultural land use, is positively and significantly correlated to both deforestation rates and deforestation levels of low income and lower middle-income countries. With respect to the underlying drivers of deforestation, evidence shows that opening to trade can reduce deforestation rates.

There are limitations to the findings and the approach. First, it aims to identify macroeconomic trends using microeconomic approaches. Growth theory traditionally assumes the existence of an aggregate production function, but there are criticisms to this approach (Banerjee and Duflo, 2005). A way to refine and improve the models would be to consider regional estimates of the production function. Further, since the data was analysed there has been updates to some of the datasets, such as the Penn World Table, now in version 10.0, and FAO's Agricultural Capitol Stock. Thus, the findings and results of Paper IV require future revision and extension. Instead of trying to focus on explaining the year-to-year changes in deforestation, perhaps a more promising approach could be to focus on the long-term cross-country differences and generate a simpler story focused on the regional differences and on the immediate drivers of deforestation.

1.6 Concluding remarks

This thesis contributes to the empirical literature evaluating the potential of collective Payments for Ecosystem Services (PES) to deliver on conservation and development outcomes under different contexts, using data from framed field experiments in Brazil, Indonesia and Peru (Paper I and II) and household survey data from Peru (Paper III). It also contributes to the literature on the drivers of deforestation, by analysing national-level drivers of deforestation using secondary data (Paper IV).

Rather than offering solid answers, some of the findings contained in the thesis invite to a lot more questions. They show that it can be hard to find a good balance between the Scylla of excessive generalizability – identifying what works 'on average' while ignoring important differences –, and the Charybdis of contextualizing impacts – a 'context' that sometimes is hard to disentangle or properly identify-. The methods and findings of this thesis are a good example of this tension.

Indeed, a repeated finding and central message of Paper I and II is 'context matters': there are important differences in relatively simple treatments across the studied sites. The findings do point out to relevant aspects that can explain these differences. Interestingly, it is structural characteristics rather than individual ones that seem to have the greatest effects in shaping game outcomes. By conducting a framed field experiment in three different countries and contextualizing the cooperation decision to a forest management situation, the thesis contributes to the evidence showing cross-country variation in experimental games.

Paper III, in turn, contains a local and highly contextualised impact evaluation, describing the actions and opinions of local actors. The paper serves as a very good example of many of the problems that conservation projects around the world face: low and insecure finance, heterogenous communities, and low impacts because of inadequate targeting. It also gives the important message that improved forest monitoring, as well as good communication between villagers and project implementers, really does matter for the perceived wellbeing of rural households.

Paper IV, in contrast, is a big generalization to attempt to answer the important question of the drivers of deforestation. The cost of such generalization is that the theoretical exercise does not lead to any straightforward policy recommendation. While it can be rewarding to identify and explain big trends, it is necessary to further refine the theoretical and empirical analysis to arrive to more solid policy recommendations. Nevertheless, the paper provides evidence that examining national deforestation through the lens of development and economic growth theories can be a fruitful endeavour to increase our understanding of the macroeconomic causes of deforestation.

2 Paper I

"While laboratory processes are simple in comparison to naturally occurring processes, they are real processes in the sense that real people participate for real and substantial profits and follow real rules in doing so. It is precisely because they are real that they are interesting." (Charles R. Plott, 1982)

Enforcement and inequality in collective PES: effectiveness, efficiency and equity implications

Julia Naime*[1,4], Arild Angelsen[1,4], Adriana Molina-Garzón[2,4], Cauê D. Carrilho[3,4], Vivi Selviana[4], Gabriela Demarchi[4,5], Amy E. Duchelle[4] and Christopher Martius[6].

[1] School of Economics and Business, Norwegian University of Life Sciences (NMBU), PO Box 5003, 1432, Ås, Norway
[2] Department of Political Science, Institute of Behavioural Science, University of Colorado Boulder, Boulder, CO, USA.
[3] Universidade de São Paulo – USP, Brazil
[4] Center for International Forestry Research (CIFOR), Bogor, Indonesia
[5] French National Research Institute for Agriculture, Food and Environment (INRAE), Montpellier, France
[6] Center for International Forestry Research (CIFOR) Germany gGmbH, Charles-de-Gaulle Strasse 5, 53113 Bonn, Germany

* Corresponding author: julia.del.carmen.naime.sanchez.henkel@nmbu.no

Abstract

Collective Payments for Ecosystem Services (PES), where forest users receive compensation conditional on group rather than individual performance, are an increasingly used policy instrument to reduce tropical deforestation. However, implementing effective, (cost) efficient and equitable (3E) collective PES is challenging because individuals have an incentive to free ride on others' conservation actions. Few comparative studies exist on how different enforcement strategies can improve collective PES performance. We conducted a framed field experiment in Brazil, Indonesia and Peru to evaluate how three different strategies to contain the free-rider problem perform in terms of the 3Es: (i) Public monitoring of individual deforestation, (ii) internal, peer-to-peer sanctions (Community enforcement) and (iii) external sanctions (Government enforcement). We also examined how inequality in wealth, framed as differences in *deforestation capacity,* affects policy performance. We find that introducing individual level sanctions can improve the effectiveness, efficiency and equity of collective PES, but there is no silver bullet that consistently improves all 3Es across country sites. Public monitoring reduced deforestation and improved the equity of the program in sites with stronger history of collective action. External enforcement provided the strongest and most robust improvement in the 3Es. While internal, peer enforcement can significantly reduce free riding, it does not improve the program's efficiency, and thus participants' earnings. Both sanctioning mechanisms failed to systematically improve the equitable distribution of benefits due to the ineffectiveness of punishments to target the largest free riders. Inequality in wealth increased group deforestation and reduced the efficiency of Community enforcement in Indonesia but had no effect in the other two country sites. Among the set of possible factors explaining differences across country sites we distinguish differences in history of collective action and land tenure systems.

Keywords: Payment for Ecosystem Services, sustainability, climate change, tropical deforestation, field experiments, common-pool resources.

2.1 Introduction

Tropical deforestation is the largest source of carbon emissions from Agriculture, Forestry and Other Land Use (AFOLU) activities (IPCC, 2019), also driving biodiversity loss (Gibson et al., 2011) and threatening the livelihoods of local communities (Angelsen et al., 2014). To meet the global climate, biodiversity and sustainable development goals, adequate policies for reducing deforestation need to be implemented at regional and local scales (Ostrom, 2010). Among the set of policy options to reduce deforestation are positive incentives (i.e., "carrots"), which aim to increase the welfare of forest users by incentivizing or rewarding their conservation activities, and disincentives (i.e., "sticks"), which aim to deter deforestation activities by punishing or increasing the cost of non-environmentally friendly behaviour (Börner et al., 2020).

Payments for Ecosystem Services (PES) programs are positive incentives that reward forest users conditional on conservation performance. They consist on voluntary agreements at the individual or group level, under which the providers agree to supply ecosystem services in exchange for payments (Wunder, 2015). PES are a commonly used tool in the efforts to reduce deforestation (Min-Venditti et al., 2017; Salzman et al., 2018) and a key component of Reducing Emissions from Deforestation and forest Degradation (REDD+) initiatives worldwide. Collective PES are characterized by assigning the payment to a group instead of an individual, based on their collective performance (Hayes et al., 2019; Pfaff et al., 2019). Collective payment is preferred when land is managed under collective ownership, when individual actions are hard to identify, or when spatial coordination of conservation activities is particularly important, such as in watershed or biodiversity management (Engel, 2016).

Although collective PES help solve the *global* collective action problem of forest conservation, they face a number of challenges to provide effective, efficient (i.e., cost-effective) and equitable outcomes (3E) (Angelsen and Wertz-Kanounnikoff, 2008) at the local level. First, they create a *local* collective action problem: the individual compensation from collective PES is only partly conditioned on individual behaviour (Hayes et al., 2019). Participants have an incentive to free ride on others' conservation actions, which can decrease the overall effectiveness of the policy as compared to an individual based PES (Kerr et al., 2012; Narloch et al., 2012; Midler et al., 2015; Gatiso et al., 2018; Hayes et al., 2019; Ngoma et al., 2020). Second, a related challenge is to balance conservation costs and benefits in a way that is equitable among participants of the program (Hayes and Murtinho, 2018; Hayes et al., 2019). Collective PES are likely to be implemented in communities with heterogenous participants in terms of household labour, capital and physical access to forests, which can in turn affect policy performance as well as exacerbate existing inequalities (Andersson et al., 2018b).

Stronger monitoring and enforcement – introducing individual "sticks" with the collective "carrots" – can help navigate these interrelated challenges. However, strong monitoring and enforcement involves additional implementation costs (Börner et al., 2014). Thus, higher program effectiveness and equity might reduce economic efficiency

(Pascual et al., 2010; Wu and Yu, 2017), yet there a few empirical evaluations of such trade-offs. In this article, we compare how different monitoring and enforcement strategies perform in terms of the 3Es in a collective PES. *Effectiveness* is the degree to which deforestation is reduced from a baseline level. *Efficiency* is the degree to which the monitoring and enforcement achieves conservation outcomes for the least cost. *Equity* is framed in terms of both a distributional and procedural dimension, and thus includes the distribution of earnings amongst PES participants as well as their fairness perceptions (Pascual et al., 2010; Loft et al., 2017; Lliso et al., 2021).

We conducted a framed field experiment (FFE) in three countries with high forest cover: Brazil, Indonesia and Peru. We compare three strategies to reduce the free rider problem: (i) Public monitoring of individual deforestation, (ii) monitoring with peer sanctions (Community enforcement) and (iii) monitoring with external sanctions (Government enforcement). We also evaluate whether inequality in wealth, framed as differences in *deforestation capacity*, affect the performance of a collective PES, as recent research suggests inequality might affect institutional performance (De Geest and Kingsley, 2021; Nockur et al., 2021). Even though a number of economic experiments have examined the effects of economic inequality on cooperation (Tavoni et al., 2011; Kingsley, 2016; De Geest and Kingsley, 2019; Hauser et al., 2019), few have tested it with actual natural resource users (Narloch et al., 2012; Vorlaufer et al., 2017; Loft et al., 2020), and none have examined the question across multiple countries.

2.2 Theoretical background

2.2.1 Reducing the free-rider problem

Collective PES programs, in which it is hard to exclude community members from the benefits of the collective payment, are similar to the common pool resource (CPR) problem; the benefit individuals receive from the group compensation is not proportional to the individual conservation actions (Martin et al., 2014; Hayes et al., 2019). To maximize own net earnings individuals can free ride by appropriating the common pool resource (i.e., deforesting), creating a negative externality on the rest of the group and reducing the collective payment.

A central strategy to reduce free riding is to increase its cost by introducing sanctions. We focus on two alternative individual sanctioning mechanisms that could be classified at the opposite sides of a governance spectrum: (i) a centralized, external sanctioning institution, and (ii) a decentralized, internal sanctioning institution in which community members sanction their peers[10]. The experimental literature indicates that in general, when faced with the threat of an external, centralized sanction, participants significantly increase cooperation (Cardenas, 2004; Rodriguez-Sickert et al., 2008; Velez et al., 2010; Lopez et al., 2012; Gelcich et al., 2013; Vollan et al., 2019). Surprisingly, the size and the

[10] Other possible sanctioning strategies between the governance spectrum include centralized, internal sanctions (see Andreoni et al. (2012) for an example), or external, collective sanctions, see Cason et al. (2013) or Suter et al. (2009).

probability of the sanctions does not greatly affect the overall effectiveness of sanctions (Cardenas, 2004; Lopez et al., 2012). This is consistent with non-experimental evidence showing how law enforcement by authorities provides effective results to reduce tropical deforestation (Busch and Ferretti-Gallon, 2017; Tacconi et al., 2019).

Likewise, experimental studies on CPRs, pioneered by Ostrom et al. (1992), show how peer punishment enhances cooperation (e.g., Kosfeld et al., 2009; Chaudhuri, 2011; Cason and Gangadharan, 2015), also in the context of collective PES (Kaczan et al., 2017). The impact of the punishment depends on the cost effectiveness of the punishment (Sutter et al., 2010; Chaudhuri, 2011), and the type of punishment – monetary or non-monetary (Masclet et al., 2003; Noussair and Tucker, 2005; Lopez et al., 2012; Pfaff et al., 2019). The experimental studies align with observational studies pointing out the capacity of communities to regulate CPR use (Ostrom, 1990; Chhatre and Agrawal, 2008; Rustagi et al., 2010). Furthermore, social, non-monetary sanctions such as the public revelation of individual decisions can increase cooperation, as it might induce guilt or shame (Masclet et al., 2003; Lopez et al., 2012; Spraggon et al., 2015; Pfaff et al., 2019). Additional factors that increase peer-punishment impact are communication (Ostrom et al., 1992; Koch et al., 2021), and previous trust and experience (Gelcich et al., 2013; Pfaff et al., 2019).

Both sanctioning strategies have potential shortfalls. External sanctions might undermine the legitimacy and liberty of participating communities, potentially crowding out motivations for cooperative behaviour (Cardenas et al., 2000; Kube and Traxler, 2011; Lopez et al., 2012). Furthermore, in many situations, external regulations and sanctioning are hard to implement, because of costly monitoring, lack of political interest, or corruption (Karsenty and Ongolo, 2012; Sundström, 2015). In turn, when communities must regulate resource use and enforce on their own, they incur monitoring and enforcement costs. If these costs are too high, they erode the benefits of more cooperation (Ostrom et al., 1992). While the efficacy of each sanctioning strategy have been evaluated in the context of homogenous populations in experimental games (see Vollan et al., 2019), there is no research evaluating how they perform relative to each other in terms of the 3E and with heterogenous populations.

2.2.2 The effect of economic inequality in the commons

It has for long been recognized that agent heterogeneity and inequality affects the level of cooperation in social dilemmas, but in ambiguous ways (Baland and Platteau, 1999; Agrawal, 2001). Broadly, three types of inequalities can affect collective action: inequality in wealth or endowments, inequality in interests or incentives, and inequality in identity (Baland and Platteau, 1996)[11]. Critical factors that determine the effect of inequality on commons outcomes include the incentive structure facing the participants

[11] Dayton-Johnson and Bardhan, (2002) further distinguish between four types of economic inequalities that are relevant in a user group: (i) inequality in wealth or income, (ii) inequalities in the sacrifices that community members make in cooperative arrangements, (iii) inequalities in the benefits they derive from public good, and (iv) inequalities from outside "exit" opportunities.

(e.g., individual endowments) and the characteristics of the public good (e.g., whether it creates positive or negative externalities and whether it offers the same returns to all participants) (Dayton-Johnson and Bardhan, 2002).

Inequality has positive effects on collective action if the wealthiest agents face stronger incentives to cooperate, for example, by receiving a larger share of the benefits from the common pool. In such cases, the elite has higher interests in collective action, and thus involve themselves more actively in setting rules and enforcing them (Baland and Platteau, 1999). Similarly, inequality in opportunity costs to a CPR (e.g., returns to the best outside option) increases cooperation, as players with more valuable external options put less pressure on the common resource (Cardenas et al., 2002). Further, an increase in wealth inequality leads to reduced deforestation when the demand for the common resource is increasing at a decreasing rate with wealth (Alix-Garcia, 2008). In this case, more inequality entails less overall deforestation because the poor reduce their deforestation more than what the wealthy increase it.

Other evidence suggests that economic heterogeneity has negative effects on the commons. For example, there is less collective action in groups with unequal landholdings (Varughese and Ostrom, 2001; Adhikari and Lovett, 2006), and more deforestation in countries with higher inequality (Koop and Tole, 2001; Ceddia, 2019). In experimental games, inequality in endowments or returns from the public good creates trade-offs between an efficient and an equitable distribution of benefits (Nikiforakis et al., 2012; Kingsley, 2016; Koch et al., 2021). Participants with higher endowments place higher value in efficiency while those with lower returns prioritize equity (Nikiforakis et al., 2012). Inequality in endowments also has negative effects on cooperation by creating distinct social identities (Martinangeli and Martinsson, 2020), decreasing levels of trust or social preferences amongst group members (Andersson and Agrawal, 2011), or reducing the positive effects of communication (Cardenas, 2003; Gangadharan et al., 2017).

In sum, the impact of inequality on the commons greatly depends on the type of inequality, the degree of inequality, and the preferences and characteristics of the group. In observational studies, the effect of economic inequality on commons outcomes is hard to identify, because different types of inequalities interact simultaneously. For example, inequality in endowment coupled with inequality in the marginal benefits from the public good can have positive effects on cooperation, but negative effects when only one type of inequality is present (Naidu, 2009; Hauser et al., 2019). Experimental methods reduce such potential sources of bias. In this paper, we use experimental data to focus on how inequality in wealth, framed as the ‘capacity to deforest’ affects participation in a collective PES.

2.3 Methods

2.3.1 Framed field experiments and the study sites

Framed field experiments (FFEs) engage actual stakeholders who have experience with the problem at hand. They recreate the decision-making situation in a controlled, hypothetical setting but with real (cash or in-kind) incentives, thus serving as a testbed of alternative real-world policy interventions (Shreedhar et al., 2020). Participants bring their own experiences and values, which increases the external validity of the results (Cardenas and Carpenter, 2008; Anderies et al., 2011; Gelcich et al., 2013; Finkbeiner et al., 2018). While FFEs never fully capture all the nuances of the actual field settings, they offer the advantage of manipulation and random assignment of treatments in a controlled setting (Ostrom, 2006), and allow for replication and direct comparison among different groups or samples.

The three sites selected for the study in Pará (Brazil), Central Kalimantan (Indonesia) and Ucayali (Peru), have characteristics that make them relevant for a comparison of the effects of a collective PES under different sanctioning institutions. At the country level, the villages share similar socioeconomic and institutional characteristics, such as drivers of deforestation and poverty levels (Sills et al., 2017). However, the country sites show differences in local reliance on forests and land tenure systems. Forests are owned communally in Peru, in Indonesia the land is owned by the state, while at the site in Brazil land is owned individually by colonist farmers. In Peru and Indonesia, households control, on average, an area of ~2.0 ha for subsistence and commercial agriculture, while in Brazil, households control on average an area of 44.8 ha of forest and 38.7 ha of agricultural land, mostly pastures.

In Brazil and Peru land tenure is in most cases considered secure, in the sense that collective and individual boundaries of properties are legally recognized. On the contrary, tenure is considered weak in Indonesia because village and households do not have legal recognition of the land they manage and forest access is based on local customary laws, which give individuals land claim when they have invested on that land (e.g., planting, clearing land) (Sills et al., 2014). Furthermore, deforestation activities by smallholders serve different economic purposes. In Indonesia, the production is mostly for subsistence consumption, while in Peru, and even more so in Brazil, it is conducted for commercial purposes. Average household deforestation is higher in Brazil (1.8 ha yr^{-1}) than in Peru (0.43 ha yr^{-1}) and Indonesia (0.04 ha yr^{-1}). Agricultural income share is higher in Peru (20.3%) than in Brazil (16.2%) and Indonesia (9.7%), while the livestock income share is much higher in the Brazilian site (47.4%) than in the Peruvian (6.4%) and Indonesian (4.7%) sites. Income inequality is highest in Brazil, but inequality in assets and land is highest in Indonesia (see Appendix A, section B4 for a detailed description of the study sites).

2.3.2 The basic experimental set-up

The FFE was implemented with 720 participants in 24 different villages between October 2019 and January 2020, equally split between the three country sites. Five experimental sessions were conducted in each village, summing up to 30 participants per village (Appendix A, section B4 for a detailed description of procedures). The average age of the participants was 44 years, and 52% of them were men.

In the experiment, a group of six forest users shared access to a forest under a collective PES, and in each round the participants simultaneously chose how many forest plots they would transform to agricultural land (croplands and pastures). Individual earnings depended on how many plots each participant had deforested and on how many forest plots were left standing once all participants had made their decisions. This framing is in contrast with previous FFEs that frame the forest management decision in terms of harvesting of forest products, and the public benefits of forest in terms of resource regrowth (Andersson et al., 2018a; Handberg and Angelsen, 2019; Ngoma et al., 2020; Palmer et al., 2020), and is considered more appropriate to how current PES programs are implemented globally and as part of REDD+.

The experiment consisted of four stages with six rounds each. In the first stage, we introduced the baseline with the collective action problem: the group benefits from the collective PES were larger than the individual gain from converting forest to agricultural land but the latter was higher than the individual benefit from the PES payment. Only the aggregate deforestation was reported back to the group. The group deforestation was perfectly monitored, and PES was fully enforced at the group level.

Inequality in wealth, or in the "capacity to deforest", was introduced by modifying the maximum number of forest plots that a participant could convert to agricultural land. It was framed in terms of household's differences in capital and labour availability for establishing agricultural plots. In half of the experimental sessions, the *Unequal* groups, three randomly chosen "low capacity" participants could deforest a maximum of four plots, and three "high capacity" participants could deforest up to eight plots. In the *Equal* groups, all participants had a "medium capacity" to deforest six plots. Thus, the aggregate deforestation capacity was the same between the Equal and Unequal groups. The experiment strictly focused on the effects of *inequality in wealth* (i.e., individual endowments) by keeping the marginal benefits of deforestation constant and equal across participants, and the same aggregate deforestation capacity across groups. Thus, the cooperation incentives were the same for every participant.

Throughout the experiment the PES payment was distributed equally among participants, as communities with collective PES often distribute the earnings based on an individual basis and on egalitarian principles, not based on individual contributions (Robinson et al., 2016; Hayes et al., 2019). Although collective payments can be subject to elite capture (Persha and Andersson, 2014; Andersson et al., 2018b), we retain the same return to be able to identify the effect of unequal wealth distribution. For the same reason, everyone received equal benefit from each plot of agricultural land. We specified that each plot was equivalent to 0.5 ha.

The stock of forestland was reset in every round, to avoid effects due to accumulated forest loss. Each plot of agricultural land was worth 10 points, while each plot of forest gave 24 points to the group, equivalent to 4 points to each player. In all sessions, each participant had a payoff table indicating his/her earnings as a function of his/her and others' decisions. Visual support was provided to explain the collective action dilemma, using a cardboard with 60 green squares. Each square represented a forest plot, and showed the group payoff of 24 points, and the individual payoff of 4 points. Whenever deforestation took place, yellow paper stickers indicating the individual payoff of 10 points replaced the green squares.

To conserve anonymity and reduce spill-overs throughout the stages, each participant was represented by a letter of the alphabet, only known to the participant and the experimenter, and the letter was changed in each stage. No communication between participants was allowed to avoid the risk of losing anonymity during the experiment by a public revelation of individual decisions or deforestation capacity. Communication was also prohibited given the interest in capturing individual motivations for conservation and sanctioning, and to avoid the emergence of different deforestation norms across groups.

2.3.3 The monitoring and enforcement treatments

Our treatments were implemented sequentially: in the second stage, after the baseline, we introduced Public monitoring. During this stage, once participants had chosen how many forest plots to deforest, the number of plots deforested by each was publicly revealed using the secret letter. The Public monitoring treatment allowed to explicitly separate the effect of two key elements of environmental governance that are often merged: monitoring and sanctioning (Andersson et al., 2014), and allowed to evaluate whether there is an effect of just increasing the amount of information available to players through announcing the individual conversion.

For the third and fourth stages, we alternated between Community enforcement and Government enforcement. The Community enforcement treatment recreated a self-enforced collective PES, in which community members themselves could choose to sanction each other to reduce free riding. The stage consisted of two steps. The first step was identical to the Public monitoring stage. In the second step, each participant chose whether or not to assign a punishment to other participants. Assigning a punishment had a cost of 10 points for the punisher but it subtracted 30 points to the punished participant. This punishment-cost ratio (3:1) follows common practice in peer punishment treatments (Chaudhuri, 2011; Vollan et al., 2019). To avoid excessive punishment, the maximum number of allowed punishments in each round was limited to three, and each punishment had to be assigned to a different participant. Information about the punisher and punished participants in each round were made public by using their secret letters. This procedure allowed retaliation and reputation building, while maintaining anonymity.

The Government enforcement treatment recreated a policy-mix scenario, in which a collective PES is implemented along with an external enforcer who randomly monitors individuals and assigns sanctions to those who deforest. During this stage, a probabilistic exposure to a third-party sanction was introduced, representing imperfect government enforcement (Cardenas et al., 2000; Velez et al., 2010). This is considered to be a better representation of the weak and costly forest enforcement that exists in most tropical forest countries (Robinson et al., 2010). The inspection probability for each participant was 1/3, and if inspected, for each plot deforested they lost 15 points. The sanction was non-deterrent as the expected benefit of deforestation was still higher than the one from conservation (i.e., it did not change the optimal strategy for a risk neutral participant). For a detailed description of the payoff functions and optimal strategies in each stage, see Appendix A (section B1).

2.3.4 Hypotheses

Evidence shows that non-monetary considerations motivate cooperative behaviour (Masclet et al., 2003; Lopez et al., 2012). At least two effects are conceivable of the Public monitoring of individual deforestation: (i) the display of non-cooperative behaviour might induce some guilt or shame of that participant (even though the revelation remains anonymous) and reduce the conversion in the following rounds; (ii) the conditional co-operators might reduce the willingness to cooperate, seeing some non-cooperative members (high converters), and thus increase deforestation.

We further expect monetary sanctions to increase cooperation, but the relative effectiveness of each enforcement strategy is difficult to predict *a priori*. Government enforcement is more likely to be more effective and efficient than Community enforcement because it imposes a norm of zero deforestation by punishing any deforestation if inspected, and it incurs no cost to participants. Community enforcement offers however, the opportunity to better target the highest free-riders (compared to random sanctioning by Government), and participants can be punished more than once. We conjecture that the effects of enforcement will differ across sites, given the difference in land tenure regimes and history of collective governance. In particular, both monitoring and peer punishment are expected to be higher in Peru and Indonesia, as compared to the Brazilian site.

The second category of hypotheses relates to the effect of inequality. Evidence from lab experiments suggests that without sanctions, inequality in individual endowment does not affect average cooperation when the aggregate endowment is the same between equal and unequal groups, as participants will move towards the non-cooperative outcome (Reuben and Riedl, 2013; Kingsley, 2016; Nockur et al., 2021). Once sanctions are introduced, participants with the highest capacity to deforest reduce their deforestation the most (Kingsley, 2016; Vollan et al., 2019). Thus, the introduction of monitoring and sanctioning should have heterogenous effects depending on the individuals' capacity to deforest. Inequality in endowments can in addition attenuate the positive effects of punishments or increase their frequency (Bernhard et al., 2006; Kingsley, 2016), increase risk taking attitudes (Payne et al., 2017), as well as reduce the preferences for internal enforcement institutions as compared to external (De Geest and Kingsley, 2019). Thus, we expect inequality in deforestation capacity to decrease the positive effects of the enforcement mechanisms, in particular efficiency.

2.3.5 Data analysis

We operationalized the 3E outcomes with three main dependent variables. To evaluate *effectiveness*, we used the group and individual deforestation levels. For *efficiency*, following Cason and Gangadharan (2015), we calculated an index based on the individuals' realized earnings (π_{it}), the self-maximizing (Nash) strategy of the baseline stage (π_{NE}) and the socially optimal payoff (π_{SO}), such that:

$$Efficiency = \frac{\pi_{it} - \pi_{NE}}{\pi_{SO} - \pi_{NE}} \tag{2.1}$$

The efficiency of each treatment compares individuals' realized payoffs π_{it} to the socially optimal outcome π_{SO}. The individual payoff includes three components: the agricultural income from forest conversion, the payment from the standing forest (the same for all group members), and the costs of received sanctions and assigned punishments during the Community and Government stages. Higher earnings indicate higher efficiency.

To measure *equity*, we considered a distributional and procedural dimension. We thus calculated a Gini coefficient in each stage (Cowell, 2011) and the perceived fairness of each enforcement strategy using a post-experimental questionnaire (Appendix A, section B2).

We used Wald tests to compare between group averages and Friedman tests or repeated measures ANOVA tests to compare within group averages- We use multilevel mixed effects linear regression models to evaluate individual level effects. We included random effects across participants and sessions in all regression models (Rabe-Hesketh and Skrondal, 2008) to control for the dependence of observations within experimental sessions and individuals. We present our main results as linear models, as they produce unbiased predictions in public good games data and their interpretation is more straightforward than probit and tobit models (Ai and Norton, 2003; Kent, 2020), and use ordered probit models as a robustness check (Moffatt, 2015).

To control for potential learning effects and temporal trends, the order of enforcement (whether Community or Government enforcement was played first), the experimental round within stages (from 1 to 6), and a dummy (from 1 to 5) indicating the order of the experimental session within a village (as there were 5 sessions in each village) were included in all the models. Likewise, to control for behavioral preferences across participants, we included variables measuring risk (Binswanger, 1981) and social preferences (Fehr et al., 2013), see Appendix A section B2 for a detailed description of elicitation methods. We also include trust as a control, given the empirical evidence indicating how trust shapes experimental outcomes (Andersson et al., 2018a; Pfaff et al., 2019). The distribution of covariates is balanced across treatments except for risk and social preferences, which are included as control in all subsequent analyses (Appendix A, section B3).

2.4 Results

2.4.1 Effectiveness

Overall, the results lend support to the hypotheses that Public monitoring works as a social sanctioning mechanism and reduces deforestation, and that introducing monetary sanctions further increases PES effectiveness. Group deforestation was high in the baseline stage: on average 15.9 and 16.8 forest plots were deforested in Equal and Unequal groups, out of a maximum of 36 (Fig. 2.1). Pairwise comparison tests indicate that Public monitoring significantly decreased group deforestation by 1.2 units in both the Equal ($p < 0.04$) and Unequal groups ($p < 0.03$), equivalent to 7.5% and 7.1% reduction respectively. In turn, Community enforcement decreased deforestation by 4.9 units or 30.8% ($p < 0.001$) in the Equal groups and by 5.7 units or 33.9% ($p < 0.001$) in the Unequal groups compared to the baseline. Government enforcement was the most effective, decreasing deforestation by 8 units or 50.3% ($p < 0.001$) in the Equal groups and by 7.5 units or 44.6% ($p < 0.001$) in the Unequal groups compared to baseline. Although group deforestation is higher in Unequal than Equal groups, the difference is not significant in any of the stages (Appendix A, Table 2.5)

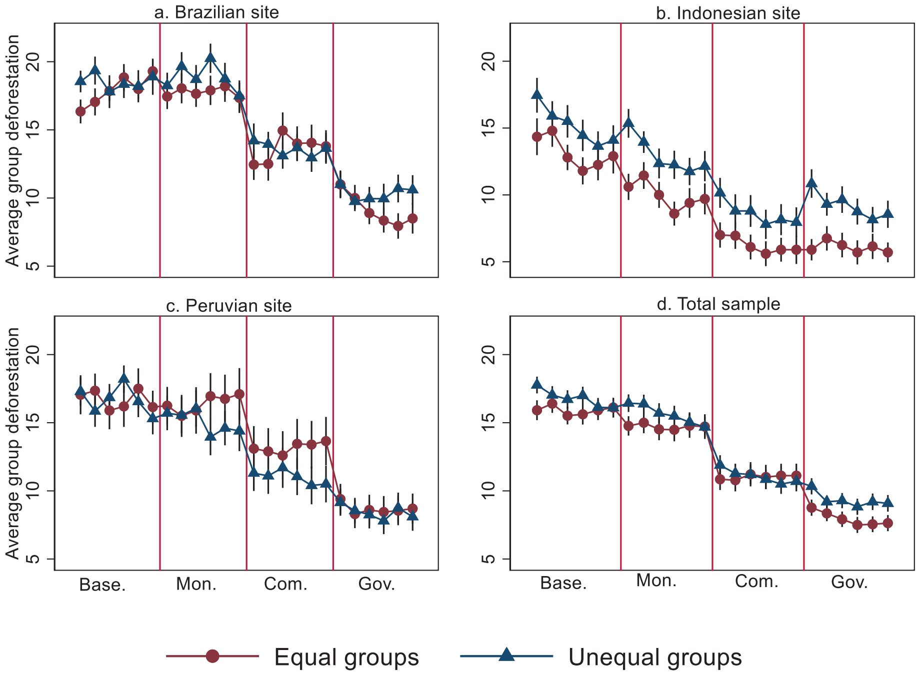


Figure 2.1 Aggregate group deforestation (number of plots) per round. Numbers at the bottom indicate the average group deforestation for each stage, in the Equal and Unequal groups. The Community and Government stages were played randomly in either rounds 13-18 or rounds 19-24. Vertical lines represent 95% confidence intervals.

There are however, important differences between the countries (Table 2.1). In Indonesia there are no differences between the treatment effects of the Community and Government enforcement (Wald test, p = 0.59), and Public monitoring had no significant effects in Brazil (Wald test, p = 0.82). Furthermore, while inequality in deforestation capacity had no effect in Brazil or Peru, it significantly increased group deforestation in Indonesia by 0.4 units or 10%. We further examined whether the effectiveness of the enforcement mechanisms is dependent on (i) the inequality treatment and (ii) the order of the enforcement. Perhaps surprisingly, we found no significant interactions with inequality (Appendix A, Table 2.8). We find, however, that the order of enforcement matters. When Community sanctions are introduced after Government enforcement, their effectiveness increases (Appendix A, Table 2.9). Thus, an important finding is that previous exposure to external enforcement increases the effectiveness of internal sanctions.

Table 2.1 Treatment effects on individual deforestation decisions, by country sites.

		(1) Total sample	**(2) Brazilian site**	**(3) Indonesian site**	**(4) Peruvian site**
Treatment					
	Monitoring	-0.20***	0.02	-0.45***	-0.16**
		(0.05)	(0.07)	(0.09)	(0.08)
	Community	-0.88***	-0.77***	-1.12***	-0.76***
		(0.07)	(0.12)	(0.11)	(0.13)
	Government	-1.29***	-1.42***	-1.09***	-1.36***
		(0.07)	(0.13)	(0.12)	(0.13)
Inequality		0.03	-0.01	0.40***	-0.35
		(0.12)	(0.14)	(0.13)	(0.24)
Constant		3.70***	3.80***	3.89***	2.69***
		(0.39)	(0.41)	(0.46)	(0.58)
Village fixed effects		Yes	Yes	Yes	Yes
Individual covariates		Yes	Yes	Yes	Yes
Observations		17280	5760	5760	5760
Log-likelihood		-30542.95	-10806.46	-9582.39	-9863.98
AIC		61181.90	21676.91	19226.79	19791.95
p-value		0.000	0.000	0.000	0.000

Note: Coefficients from multilevel mixed effects linear models of deforestation, with random effects at the experimental session and individual level. Clustered standard errors at the experimental session level in parenthesis. P-values * $p < 0.1$, ** $p < 0.05$, *** $p<0.01$.

Separating the effects by participant type (i.e., by their deforestation capacity) reveals that participants with a high (low) deforestation capacity deforested more (less) than their medium-capacity counterparts (Table 2.2). Importantly, there are heterogenous responses to treatment depending on the participant type. For instance, the Public monitoring effect in Peru is dominated by the response of wealthy (i.e., high deforestation capacity) participants (Table 2.2, column 4). In general, wealthy participants responded more to the Community and Government enforcement, while the behavioural response from participants with low deforestation capacity was in general weaker. As a result, there were no significant differences in predicted deforestation levels among participant types during the Community and Government enforcement stages in any country (Fig. 2.2). In other words, the introduction of sanctions equalized individual deforestation levels.

We further examined the proportion of forest plots deforested from the maximum allowed (instead of the absolute deforestation levels) and found no significance in the interaction terms (Appendix A, Table 2.10). Thus, the heterogenous effects by participant type manifest in absolute changes in deforestation, not in relative changes. Furthermore, participants in Indonesia with low deforestation capacity converted a higher proportion than their medium-capacity counterparts, which explains why there are no significant differences in absolute deforestation levels between the two groups levels (Table 2.2).

Table 2.2 Treatment effects on individual deforestation (effectiveness).

	(1) Total sample	(2) Brazilian site	(3) Indonesian site	(4) Peruvian site
Treatment				
Monitoring	-0.20***	-0.02	-0.53***	-0.05
	(0.07)	(0.09)	(0.13)	(0.11)
Community	-0.82***	-0.71***	-1.15***	-0.58***
	(0.10)	(0.17)	(0.16)	(0.19)
Government	-1.33***	-1.46***	-1.18***	-1.34***
	(0.10)	(0.15)	(0.17)	(0.20)
Deforestation capacity				
Low capacity	-0.52***	-0.76***	-0.17	-0.70**
	(0.13)	(0.20)	(0.18)	(0.28)
High capacity	0.60***	0.68***	0.78***	0.29
	(0.16)	(0.23)	(0.27)	(0.30)
Interaction terms				
Monitoring*Low	0.13	0.19	0.32*	-0.11
	(0.09)	(0.12)	(0.17)	(0.14)
Community*Low	0.25*	0.40*	0.32	0.04
	(0.13)	(0.22)	(0.20)	(0.22)
Government*Low	0.48***	0.54**	0.53**	0.37
	(0.14)	(0.23)	(0.21)	(0.26)
Monitoring*High	-0.12	-0.04	0.00	-0.34*
	(0.13)	(0.20)	(0.25)	(0.19)
Community*High	-0.53***	-0.62**	-0.20	-0.76**
	(0.17)	(0.29)	(0.29)	(0.32)
Government*High	-0.32*	-0.35	-0.16	-0.45
	(0.19)	(0.34)	(0.34)	(0.30)
Constant	3.60***	3.70***	3.81***	2.64***
	(0.39)	(0.42)	(0.50)	(0.57)
Village fixed effects	Yes	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes	Yes
Observations	17280	5760	5760	5760
Log-likelihood	-30448.05	-10765.64	-9557.87	-9821.92
AIC	61006.11	21609.28	19191.74	19721.83
p-value	0.00	0.00	0.00	0.00

Note: Coefficients from multilevel mixed effects linear models of deforestation, with random effects at the experimental session and individual level. Clustered standard errors at the experimental session level in parentheses. P-values * $p < 0.1$, ** $p < 0.05$, *** $p<0.01$.

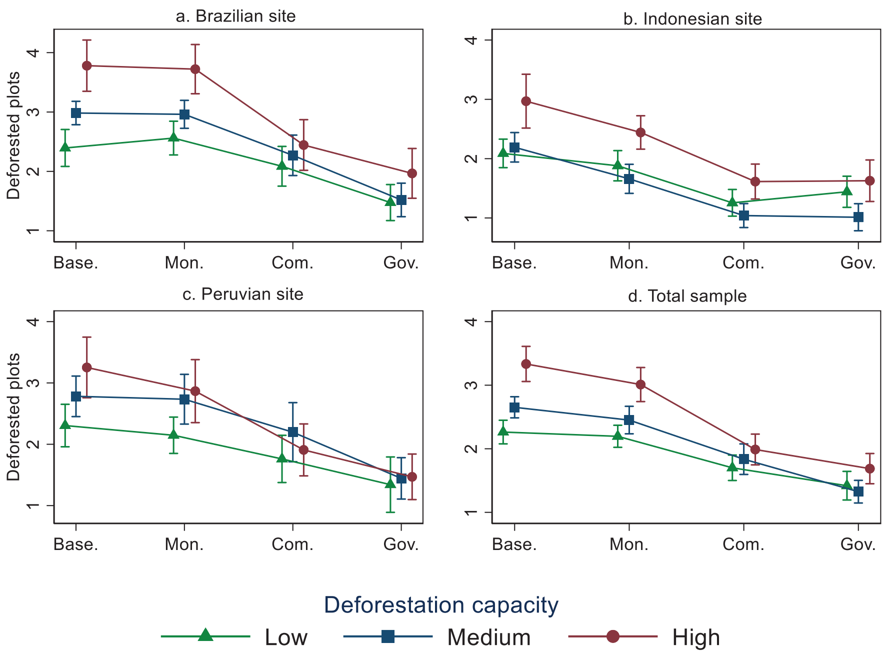


Figure 2.2. Predicted deforestation depending on participant's deforestation capacity, by treatment and country site. Vertical lines represent 95% confidence intervals.

2.4.2 Efficiency

Recall that the efficiency index is individuals' realized payoffs relative to the socially optimal outcome, cf. Eq. (1). Public monitoring of individual deforestation increased efficiency in Indonesia and Peru. Government enforcement was the most efficient treatment in both Equal and Unequal groups, and in all countries (Table 2.3). This result is not only contingent on the fact that Government enforcement had no costs to participants during the experiment: Government sanctions would have to be at least four times more costly than Community sanctions per individual monitored to reverse this finding (Appendix A, Fig. 2.5). Community enforcement, on the other hand, did not increase efficiency compared to the baseline stage, in any of the country sites (Table 2.3). Thus, the benefits of the disciplining effect of peer punishment were not sufficient to outweigh its cost. Moreover, in Unequal groups in Indonesia and overall, Community enforcement *decreased* efficiency participants' earnings (Table 3, columns 2 and 6). The lower efficiency observed in the Unequal groups during the Community stage is explained by the higher frequency of costly punishment in Unequal groups (16.9 per session on average) as compared to the Equal groups (11.7 per session on average), a statistically significant difference (Appendix A, Table 2.11).

Table 2.3 Treatment effects on efficiency

	Total sample		Brazilian site		Indonesian site		Peruvian site	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Monitoring	0.03^{***}	0.03^{***}	-0.00	0.00	0.07^{***}	0.09^{***}	0.03^{**}	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
Community	-0.05^{***}	-0.02	-0.03	-0.01	-0.11^{***}	-0.04	-0.02	-0.01
	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(0.03)
Government	0.13^{***}	0.14^{***}	0.14^{***}	0.16^{***}	0.10^{***}	0.14^{***}	0.14^{***}	0.14^{***}
	(0.01)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)
Inequality	-0.04^{*}	-0.02	-0.04	-0.02	-0.12^{***}	-0.06^{**}	0.02	0.02
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Interaction terms								
Monitoring		-0.00		-0.02		-0.03		0.04
*Inequality		(0.02)		(0.02)		(0.03)		(0.02)
Community		-0.06^{*}		-0.03		-0.14^{*}		-0.02
*Inequality		(0.04)		(0.06)		(0.08)		(0.05)
Government		-0.03		-0.04		-0.07		-0.00
*Inequality		(0.03)		(0.05)		(0.05)		(0.05)
Constant	0.35^{***}	0.34^{***}	0.37^{***}	0.36^{***}	0.20^{**}	0.17^{**}	0.54^{***}	0.54^{***}
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)	(0.10)
Village fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind. covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17280	17280	5760	5760	5760	5760	5760	5760
Log likelihood	-1317.5	-1295.5	-530.9	-528.9	-642.7	-613.7	69.4	75.1
AIC	2731.1	2693.2	1125.8	1127.9	1347.5	1295.5	-74.8	-80.2
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: Coefficients from multilevel mixed effects linear models of deforestation, with random effects at the experimental session and individual level. Clustered standard errors at the experimental session level in parentheses. P-values * $p < 0.1$, ** $p < 0.05$, *** $p<0.01$.

2.4.3 Equity and fairness

Inequality (i.e., Gini coefficient) decreased with the introduction of Public monitoring and Government enforcement, but not with the introduction of Community enforcement (Table 2.4). Both Public Monitoring and Government enforcement decreased inequality in earnings compared to the baseline stage in the Equal groups, while in the Unequal groups only Public monitoring had a significant effect in reducing inequality in earnings (Fig. 2.3). Considering individual countries, in Brazil none of the enforcement strategies reduced inequalities in the earnings. In Peru only Public monitoring had an effect. In Indonesia, Community enforcement *increased* inequality when it was implemented in Unequal groups, and in Equal groups both Government enforcement and Public monitoring reduced inequality (Fig. 2.3).

Table 2.4 Average Gini coefficient[12] in the Equal and Unequal groups, by stage. Standard deviations in parenthesis. Friedman tests indicate significant differences between the Gini coefficients of each stage in the Equal (p=0.007) and Unequal groups (p<0.001).

Gini coefficient	**Baseline**	**Monitoring**	**Community**	**Government**
Equal groups	0.041 (0.01)	0.038 (0.02)	0.043 (0.03)	0.034 (0.02)
Unequal groups	0.045 (0.02)	0.041 (0.02)	0.052 (0.03)	0.040 (0.02)

Why did the treatments not reduce inequalities significantly, despite deforestation rates being equalized across participant types? When considering the Gini coefficient of earnings without including the punishment costs, there are significant reductions in inequalities (Table 2.13 and Fig. 2.6, Appendix A). Thus, it is the punishment behaviour during the Community monitoring, as well as the random nature of sanctioning from the part of Government which inhibits positive distributional effects of enforcement.

Despite the low positive distributional effect, Government enforcement was perceived as fairer than Community enforcement. Half (51.1%) of the participants thought that Government enforcement was fairer than Community enforcement, while 24.6% favored Community over Government enforcement. The rest of the participants considered that both enforcements were equally fair (21.3%) or that neither institutional arrangement was fair (3%). In Peru participants were more likely to mention that both types of enforcement were equally fair (41%), while in Indonesia and Brazil most participants thought Government enforcement was fairer, with 64 % and 54 % of the participants, respectively. The probability of choosing either Government or Community enforcement as fairer was independent of being a participant with high, medium or low deforestation capacity (see Table 2.14, Appendix A).

12 Gini coefficients are in general low because the collective benefits were relatively large. We deliberately chose to have a high base payment to participants and to reduce the absolute differences in the final payments, for ethical reasons.

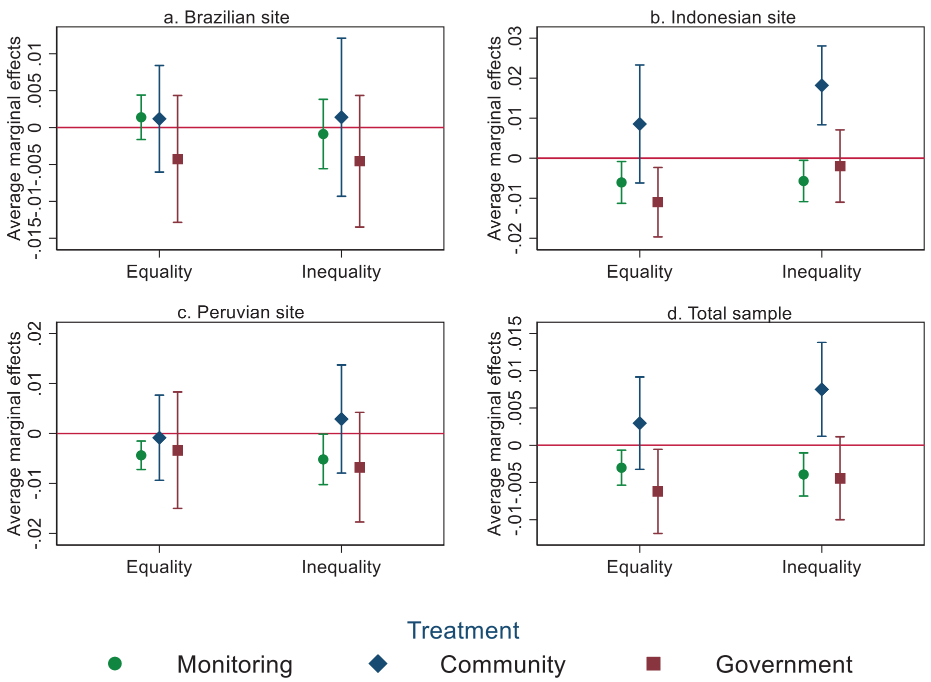


Figure 2.3 Average marginal treatment effects of Public monitoring, Community and Government enforcement on the Gini coefficient, for Equal and Unequal groups and by country. See SI (Table S8) for full model specification and regression results. Vertical lines represent 95% confidence intervals of the coefficients.

2.5 Discussion

Collective payments for forest conservation create a local collective action problem, as individual forest users have incentives to free ride on others' conservation actions. Introducing individual level sanctions can improve the effectiveness, efficiency and equity of collective PES, but a main finding is that there is no strategy that simultaneously and consistently improves the 3E outcomes across country sites and inequality contexts.

2.5.1 Reducing the free-rider problem

Public monitoring of individual deforestation had a positive, albeit modest effect on group deforestation. This is consistent with studies showing that monitoring activities can increase PES effectiveness (Martin et al., 2014) and forest protection in general (Slough et al., 2021a), but also that they are far from being sufficient to ensure perfect compliance (Wunder et al., 2018). In our study, the effect was significant only in the country sites which have history of collective action in terms of forest management and rule setting (Peruvian and Indonesian sites). This suggests that previous experience with collective agreements is an essential ingredient for getting a positive conservation impact of individual monitoring. The experimental literature has also demonstrated how previous communication or successful cooperation positively influences collective outcomes (Gangadharan et al., 2017; Rodriguez et al., 2019). While in our experiment the individual monitoring was anonymised, non-anonymised reporting, where the identity of the individuals is revealed, could have yielded even stronger effects. For example, public disclosure has stronger effects when non-cooperating individuals are singled out (Spraggon et al., 2015).

Government enforcement is the most robust policy to increase the effectiveness and efficiency of the collective PES and was effective in all country sites and inequality contexts. Introducing external sanctions allows to coordinate on particular norms that can serve as focal-points (Nikiforakis et al., 2012). This increases cooperation and also explains why Community enforcement became more effective when introduced after the Government sanction. We also show that the ineffective (i.e., random) targeting of largest free riders inhibits the positive distributional effects of enforcement. Thus, accurately identifying the largest free riders is necessary to strengthen the positive equity effect of external enforcement. An impartial, strong external enforcement might be difficult to implement in situations of weak governance and corruption, where private interests or lack of funding might conflict with the provision of the public goods (Karsenty and Ongolo, 2012; Sundström, 2015). This is still a major challenge for effective environmental regulation. Nonetheless, most participants perceived Government enforcement as being fair, which indicates that effectiveness and efficiency considerations do not contradict equity and fairness ones. Emphasizing the potential win-win outcomes of external sanctions is particularly important considering that the enforcement and sanctioning of PES non-compliance often lacks political support (Wunder et al., 2018).

Community enforcement can deliver on conservation outcomes but at a significant cost to community members. Results from the Indonesian site show that compared to the baseline stage, introducing costly peer punishment creates significant trade-offs between effectiveness on the one hand, and efficiency and equity on the other. Community enforcement costs could be reduced if collective PES implementers facilitate communication and increase social capital amongst PES participants. A large body of experimental evidence has shown the positive effects of communication on cooperation, which tend to be greater in homogenous rather than heterogenous groups (Hackett et al., 1994; Cardenas et al., 2002; Chaudhuri, 2011; Tavoni et al., 2011; Gangadharan et al., 2017). Non-experimental studies suggest stakeholder involvement and external support from intermediaries such as NGOs facilitate participation and cooperation in PES in general (Pham et al., 2010; Murtinho and Hayes, 2017; Izquierdo-Tort et al., 2021), and can reduce elite capture (Persha and Andersson, 2014). Given that strong community governance remains a major challenge (Dokken et al., 2014; Murtinho and Hayes, 2017), our study highlights the need to guarantee that communities have an arena to discuss strategies and define their monitoring and sanctioning rules in the implementation of collective PES.

2.5.2 The effect of inequality

Our study provides new evidence of how wealth inequality, understood as differences in the capacity to engage in deforestation, impacts the effectiveness and efficiency of environmental regulations. The effect of wealth inequality cannot, however, be generalized across study sites: it was only significant in Indonesia, where it both increased deforestation as well as the frequency of peer punishment. Three factors potentially explaining the strong inequality effect in the Indonesia site are: higher inequality in landholdings compared to the other two sites, lower level of tenure security, and stronger customary rules of forest management. These factors can also explain why there were no differences in the effectiveness of external and internal enforcement in this country site, coinciding with a similar experiment conducted in Namibia (Vollan et al., 2019). While the impact of inequality seems to be largely shaped by context (i.e., country site), future research could examine how this effect is mediated by factors such as levels of trust and social preferences amongst participants. The heterogenous findings across sites highlight the importance of considering different populations in experimental studies.

A result generalizable across country sites is that wealthy participants with high deforestation capacity tended to be more responsive to the threat of sanctions than their poorer counterparts. This result is particularly interesting considering that all participants faced the same incentives to cooperate and the same sanctioning costs. The lower responsiveness of poorer participants to sanctioning is consistent with being more averse to disadvantageous inequality than to advantageous inequality (Fehr and Schmidt, 1999). Evaluations of collective PES also show that it is wealthier residents who are more likely to change their behaviours (Hayes et al., 2017).

2.5.3 Policy implications

Two important considerations for the external validity and policy implications of our results are first, that the endowment inequality was created exogenously. Different results could be expected with endogenous inequality (i.e., with a real effort task), as the origin of wealth differences affects fairness perceptions (Almås et al., 2010). Second, the experiment simulated a best-case scenario of perfect and costless monitoring conditions: PES was perfectly monitored, and everyone could observe others' deforestation and could punish all players at the same cost (Community stage) or with the same probability (Government stage).

Arguably, conditions in the field are different; it might be costly to track individual deforestation, or power relations can modify enforcement costs amongst community members. Experimental evidence shows that external enforcement maintains strong effects even with lower sanctioning probabilities than in this study (Lopez et al., 2012; Vollan et al., 2019), or when the sanctions are provided at the collective rather than individual level (Cason and Gangadharan, 2013). On the other hand, under imperfect monitoring, the effectiveness and efficiency of peer punishments decreases (Grechenig et al., 2010; Boosey and Isaac, 2016; Shreedhar et al., 2020), as do the acceptability and preference for a decentralized institution (De Geest and Kingsley, 2019). These findings – along with our results – point to the advantages of external enforcement as compared to internal enforcement mechanisms when implementing collective PES. Given the known positive effects of community monitoring in the management of common-pool resources (Buntaine and Daniels, 2020; Slough et al., 2021b), a combination of bottom-up monitoring with top-level enforcement could be another promising strategy to increase individual compliance in collective agreements. Yet, it could potentially decrease the economic efficiency (earnings) as the PES participants incur the monitoring costs.

2.6 Conclusion

Collective payments are a promising conservation policy to reduce global deforestation, but their effectiveness is jeopardized by the fact that they entail incentives for individual free riding. As collective PES gain traction, policy makers and practitioners should consider strategies that can help solve the free-riding problem intrinsic to such payments and thus deliver effective, efficiency and equitable (3E) outcomes. Our study is the first to show the implications of different strategies to limit free riding in collective payments on the 3E outcomes. Compared to a situation of collective PES without any individual monitoring and enforcement, we show that introducing monitoring and enforcement allows to significantly increase the conservation benefits of collective PES, although the impacts vary greatly by context.

Public monitoring of individual decisions has limited effectiveness as compared to the introduction of monetary sanctions, and a significant effect is only observed in sites with a stronger history of collective action. Community enforcement (internal, peer-to-peer sanction) increases effectiveness but can reduce the efficiency and equity of collective PES, especially when implemented in communities with unequal access to resources. We find important variations in impacts; for example, in Indonesia the reduction in deforestation from Community enforcement is higher than in the other two sites, and inequality in the access to forest resources significantly increases group deforestation. However, across the sites, external, Government enforcement provides the strongest and most robust results in terms of effectiveness and efficiency outcomes. We further show that a costly punishment that does not effectively target free riders hampers the positive distributional effects of both enforcement strategies.

Finally, we find that implementing collective PES in situations with inequality in wealth can have negative effects on conservation and exacerbate the trade-offs between effectiveness, efficiency and equity outcomes. In addition to individual free riding, a challenge in designing and implementing PES is to manage such trade-offs, and our results suggest that these are particularly pronounced - and thus PES implementation more challenging - in contexts with unequal forest access. The results are relevant for both collective PES schemes as well as group-based incentive schemes in general.

Acknowledgements

This research is part of CIFOR's Global Comparative Study on REDD+ (www.cifor.org/gcs). The funding partners that have supported this research include the Norwegian Agency for Development Cooperation (Norad), the Australian Department of Foreign Affairs and Trade (DFAT), the European Commission (EC), the International Climate Initiative (IKI) of the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU), the United Kingdom Department for International Development (UKAID), and the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA), with financial support from the donors contributing to the CGIAR Fund. We thank Krister Andersson, Jan Börner, Federico Cammelli, Nathan Cook, Santiago Izquierdo-Tort and Sven Wunder for valuable comments to an earlier version of our manuscript. All participants in the study sites in Brazil, Indonesia and Peru are gratefully acknowledged.

Appendix A

This section presents supplementary results, in particular group-level comparisons (section A1), individual treatment effects (A2), a comparison of efficiency levels between Community and Government enforcement (A3), the analysis of individual fairness perceptions and treatment effects on inequality (Gini) index (A4), and supplementary regressions using tobit and probit models (A5).

A.1. Group-level comparisons.

Table *2.5* presents the average group deforestation for each stage, separating between the Equal and Unequal groups. The group means approximate a normal distribution (Fig. 2.4). The p-values of t-tests indicate that there are no significant differences in average deforestation between the Equal and Unequal groups for each stage when the unit of analysis is the experimental session (N=120). However, there are significant differences between the Equal and Unequal groups when the unit of analysis is the experimental round, in particular during the Baseline and Government enforcement stages (N=720). These results suggest the effect of inequality might be too small to be detectable with our sample size.

Stage	**Average group deforestation**		**Differences in means (Equal-unequal)**	**p-value of t-test (session level) (N=120)**	**p-value of t-test (round level) (N=720)**
	Equal	**Unequal**			
Baseline	15.9 (0.15)	16.8 (0.14)	-0.9	0.42	0.07
Monitoring	14.7 (0.17)	15.6 (0.15)	-0.9	0.48	0.10
Community	11.0 (0.18)	11.1 (0.15)	-0.1	0.96	0.93
Government	7.9 (0.12)	9.3 (0.12)	-1.4	0.14	0.001
All stages	**12.4 (0.16)**	**13.2 (0.13)**	**-0.8**	**0.16**	**0.004**

Table 2.5. Average group deforestation by stage and inequality context.
Two-sided t-tests between the Equal and Unequal groups. Standard errors indicated in parenthesis.

Table 2.6 presents the results of pairwise comparisons of average group deforestation between treatments (Public monitoring, Community enforcement and Government enforcement), derived from repeated measures ANOVA and separating between Equal and Unequal groups. They indicate significant differences in group deforestation as compared to the baseline, and significant differences between the treatments in both the Equal and Unequal sessions.

Table 2.7 presents the same analysis using group efficiency as dependent variable. In the Equal groups, there are significant differences in average efficiency for all treatments as compared to the baseline stage, except for the Community stage. Furthermore, in the Unequal groups, there are significant *reductions* in efficiency during the Community enforcement stage (marked bold) compared to the baseline stage and the monitoring stage.

Table 2.6. Difference in average group deforestation between stages. Difference in the Equal groups are indicated in the blue cells, while differences in the Unequal groups are indicated in the red cells (N=240). P-values in parenthesis correspond to pairwise comparison of repeated ANOVA tests.

Difference in deforestation	Baseline	Monitoring	Community	Government
Baseline		-1.2 (p<0.03)	-5.7 (p<0.001)	-7.5 (p<0.001)
Monitoring	-1.2 (p<0.04)		-4.5 (p<0.001)	-6.3 (p<0.001)
Community	-4.9 (p <0.001)	-3.7 (p <0.001)		-1.8 (p=0.007)
Government	-8.0 (p <0.001)	-6.8 (p <0.001)	-3.1 (p <0.001)	

Table 2.7. Difference in average efficiency between stages. Difference in the Equal groups are indicated in the blue cells, while differences in the Unequal groups are indicated in the red cells. Bold cells indicate negative or insignificant differences in efficiency (N=240). P-values in parenthesis correspond to pairwise comparison of repeated ANOVA tests.

Difference in efficiency	Baseline	Monitoring	Community	Government
Baseline		**0.03 (p=0.2)**	**-0.08 (p=0.001)**	0.11 (p<0.001)
Monitoring	0.03 (p=0.10)		**-0.11 (p<0.01)**	-0.8 (p=0.001)
Community	**-0.02 (p=0.35)**	0.05 (p =0.01)		0.19 (p<0.001)
Government	0.15 (p <0.001)	0.11 (p <0.001)	0.16 (p <0.001)	

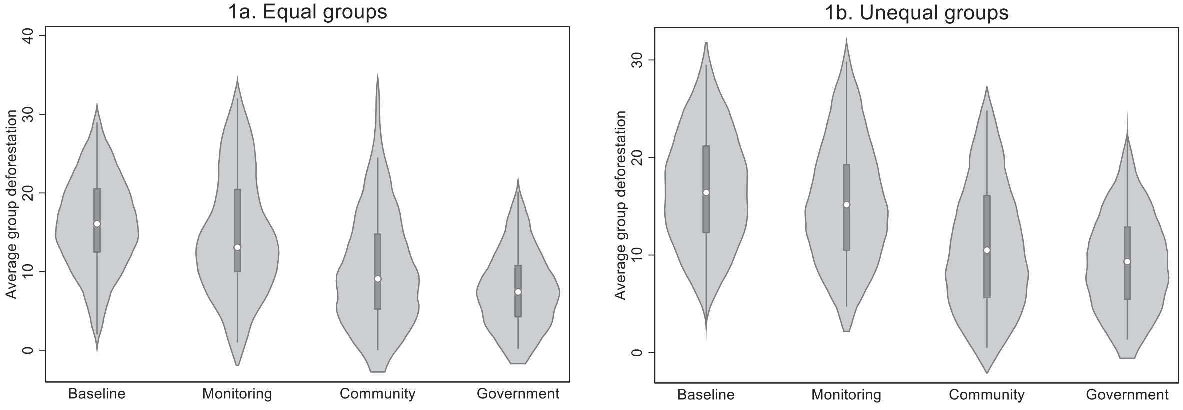


Figure 2.4. Violin plots of average group deforestation during each treatment, in Equal groups (panel 1a) and Unequal groups (panel 1b). The white marker indicates the median deforestation, the shaded area the kernel density distribution and the box the interquartile range.

A2. Individual treatment effects

Table 2.8 presents the treatments effects and the interaction effect of inequality. They indicate that on average, the existence of inequality does not influence deforestation levels, except for the Indonesian case. Interaction terms are not significant, thus monitoring and enforcement effects are not mediated by the existence of inequality.

Table 2.9 indicates the interaction effects of the order of enforcement on the effectiveness of the monitoring and sanctioning strategies. Introducing first the Government enforcement increases the effectiveness of the Community enforcement, for the whole sample and in all countries. This lends support to the hypothesis that Government enforcement acts as a focal point and establishes stronger norms of cooperation during Community enforcement.

Table 2.10 presents the analysis using the proportion of plots deforested from the maximum allowed. We find no significant effects in the interaction terms for the whole sample (column 1). However, participants with low capacity to deforest converted a higher proportion than their medium capacity counterparts.

Table 2.8. Multilevel linear mixed-effects models of individual deforestation with random effects at the individual and experimental session level, with treatment interaction.

Dependent variable: plots deforested.	(1) Total sample	(2) Brazilian site	(3) Indonesian site	(4) Peruvian site
Monitoring	-0.200***	-0.022	-0.532***	-0.047
	(0.070)	(0.085)	(0.132)	(0.106)
Community	-0.816***	-0.712***	-1.151***	-0.585***
	(0.104)	(0.171)	(0.158)	(0.191)
Government	-1.327***	-1.464***	-1.179***	-1.337***
	(0.101)	(0.151)	(0.171)	(0.197)
Inequality	0.041	-0.022	0.318**	-0.193
	(0.122)	(0.169)	(0.157)	(0.250)
Monitoring*Inequality	0.005	0.076	0.164	-0.225
	(0.097)	(0.137)	(0.181)	(0.148)
Community*Inequality	-0.138	-0.110	0.057	-0.360
	(0.138)	(0.235)	(0.211)	(0.251)
Government*Inequality	0.082	0.097	0.185	-0.036
	(0.146)	(0.266)	(0.235)	(0.250)
Constant	3.810***	4.117***	3.814***	2.768***
	(0.406)	(0.459)	(0.567)	(0.636)
Village fixed effects	Yes	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes	Yes
Observations	17280	5760	5760	5760
p-value	0.000	0.000	0.000	0.000

Note: Clustered standard errors at the experimental session level in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p<0.01$.

Table 2.9 Multilevel linear mixed-effects models of individual deforestation with random effects at the individual and experimental session level, with interactions with the order of enforcement.

	(1) Total sample	(2) Brazilian site	(3) Indonesian site	(4) Peruvian site
Monitoring	-0.21***	-0.01	-0.35**	-0.28**
	(0.07)	(0.09)	(0.14)	(0.12)
Community	-0.67***	-0.58***	-0.95***	-0.48**
	(0.10)	(0.13)	(0.14)	(0.20)
Government	-1.32***	-1.49***	-1.04***	-1.43***
	(0.11)	(0.19)	(0.17)	(0.20)
Inequality	0.03	-0.01	0.42***	-0.35
	(0.12)	(0.14)	(0.12)	(0.24)
Effect of Order of enforcement				
Government played first (dummy)	-0.07	-0.04	0.04	-0.18
	(0.12)	(0.18)	(0.15)	(0.23)
Interaction terms				
Monitoring* Government played first	0.03	0.06	-0.21	0.24
	(0.10)	(0.14)	(0.18)	(0.15)
Community* Government played first	-0.43***	-0.38*	-0.35*	-0.58**
	(0.13)	(0.23)	(0.20)	(0.24)
Government* Government played first	0.07	0.16	-0.09	0.15
	(0.15)	(0.27)	(0.24)	(0.25)
Constant	3.77***	4.09***	3.68***	2.82***
	(0.41)	(0.47)	(0.54)	(0.66)
Village fixed effects	Yes	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes	Yes
Observations	17280	5760	5760	5760
p-value	0.00	0.00	0.00	0.00

Note: Clustered standard errors at the experimental session level in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.10. Multilevel tobit mixed-effects models left censored models of individual relative deforestation, with random effects at the individual and experimental session level.

Dependent variable: proportion of plots deforestation from maximum allowed	(1) Total sample	(2) Brazilian site	(3) Indonesian site	(4) Peruvian site
Monitoring	-0.05***	-0.01	-0.13***	-0.02
	(0.02)	(0.02)	(0.03)	(0.02)
Community	-0.19***	-0.15***	-0.30***	-0.14***
	(0.03)	(0.04)	(0.05)	(0.05)
Government	-0.33***	-0.34***	-0.31***	-0.33***
	(0.03)	(0.04)	(0.05)	(0.05)
Low capacity	0.11***	0.08	0.16***	0.09
	(0.03)	(0.05)	(0.05)	(0.06)
High capacity	-0.03	-0.05	0.01	-0.06
	(0.03)	(0.04)	(0.05)	(0.06)
Monitoring*Low	0.03	0.05*	0.06	-0.03
	(0.02)	(0.03)	(0.05)	(0.03)
Monitoring*High	0.00	0.00	0.06	-0.04
	(0.02)	(0.03)	(0.04)	(0.03)
Community*Low	0.01	0.05	0.01	-0.03
	(0.04)	(0.06)	(0.07)	(0.06)
Community*High	-0.03	-0.06	0.06	-0.08
	(0.03)	(0.05)	(0.06)	(0.06)
Government*Low	0.04	0.03	0.08	0.02
	(0.04)	(0.07)	(0.06)	(0.07)
Government*High	0.04	0.03	0.07	0.01
	(0.04)	(0.06)	(0.06)	(0.06)
Constant	0.66***	0.63***	0.78***	0.38***
	(0.09)	(0.07)	(0.12)	(0.14)
Village fixed effects	Yes	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes	Yes
Observations	17280	5760	5760	5760
p-value	0.00	0.00	0.00	0.00

Note: Clustered standard errors at the experimental session level in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

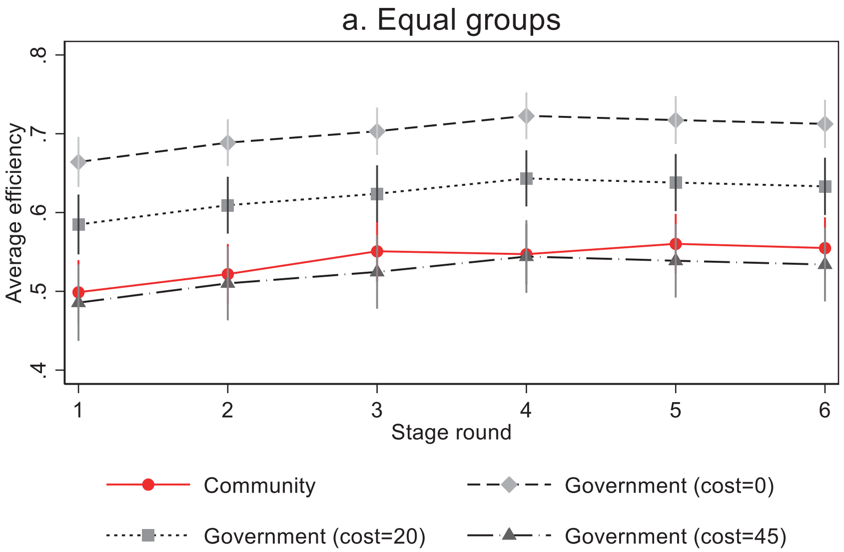
A3. Comparison of efficiency levels between Government and Community

In the main results, the analysis of efficiency considered only Community enforcement costs. By design, Government enforcement was costless to the group. However, in real-world scenarios government enforcement is costly to society. To further compare the efficiency of Community enforcement and Government enforcement, a scenario where Government enforcement is costly was recreated. Introducing a unitary cost of Government enforcement of 20 points or of 45 points per monitored individual (recalling that the costs of a community sanction is 10 points per punishment) indicates that with a cost of 45 points per individual monitored, the efficiency of Government enforcement is as low as in Community enforcement, in both the Equal and Unequal groups. Figure 2.5 presents the results of such comparison. We estimated government costs to be per participant monitored; if we considered the cost per participant actually being punished, the efficiency-equivalent government costs as compared to Community enforcement would have been even higher.

Table 2.11 compares the frequency of punishments in the Equal and Unequal groups. There are significant differences in overall punishment levels, with a total of 706 punishments in the Equal sessions and 1015 punishment in the Unequal sessions. Punishment decisions can be classified as antisocial or prosocial. Antisocial punishment occurs when participants who deforest less than group average are punished, and prosocial punishment occurs when participants who deforest more than group average are punished. Players who engaged in prosocial costly punishment potentially generated public benefits, as they discouraged future deforestation. Thus, they contributed to a second order cooperation dilemma (i.e., creating the public good of rule enforcement) (Rustagi et al., 2010). The share of pro-social punishments was slightly higher in the Equal sessions (66.5%) than in the Unequal sessions (62.9%).

Table 2.11. Punishment decisions of the Equal and Unequal sessions, separating between pro-social and anti-social punishments. Standard errors in parenthesis.

Average punishment per round	**Equal sessions**	**Unequal sessions**	**p-value of Mann–Whitney U-test (N=120)**
All punishment	11.7 (1.43)	16.92 (1.90)	0.06
Pro-social punishments	7.78 (0.96)	10.65 (1.17)	0.11
Antisocial punishments	3.92 (0.58)	6.27 (0.89)	0.18



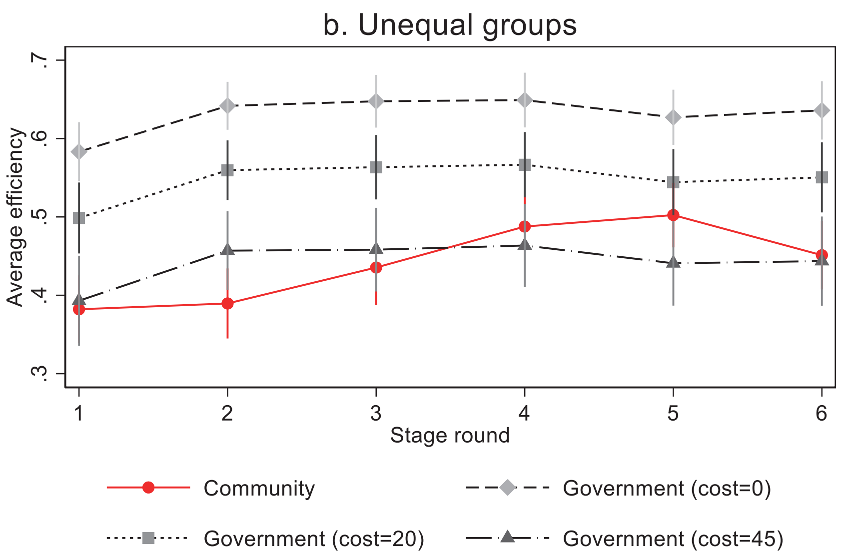


Figure 2.5. Average efficiency per round during Community and Government enforcement, in Equal groups (panel a) and Unequal groups (panel b). Panels indicate the observed efficiency of Community and Government enforcement as designed during the experiment, as well as two additional scenarios where cost of Government enforcement is 20 or 45 points per monitored individual. Vertical lines indicate 95% confidence intervals.

A4. Fairness perceptions and effect on Gini coefficients

Table 2.12 indicates the treatment effects on the Gini coefficient, for the total sample as well as for individual country sites (columns 2-4). Public monitoring had an effect in reducing inequality in Indonesia as well as Peru. There was no significant effect of any treatment in Brazil.

Table 2.13 demonstrates that when considering the earnings without the sanctioning costs of either Community or Government enforcement, the monitoring and enforcement treatments do have a significant effect in reducing inequalities within groups. Figure 2.6 shows the marginal effects in each country.

Table 2.12. Multilevel mixed models of the effect of treatments on the Gini coefficient, calculated using sanctioning costs.

	(1) Total sample	(2) Brazilian site	(3) Indonesian site	(4) Peruvian site
Monitoring	-0.003**	0.001	-0.006**	-0.004***
	(0.00)	(0.00)	(0.00)	(0.00)
Community	0.003	0.001	0.009	-0.001
	(0.00)	(0.00)	(0.01)	(0.00)
Government	-0.006**	-0.004	-0.011**	-0.003
	(0.00)	(0.00)	(0.00)	(0.01)
Inequality	0.002	0.004	0.004	-0.002
	(0.00)	(0.00)	(0.00)	(0.00)
Monitoring*Inequality	-0.001	-0.002	0.000	-0.001
	(0.00)	(0.00)	(0.00)	(0.00)
Community*Inequality	0.005	0.000	0.010	0.004
	(0.00)	(0.01)	(0.01)	(0.01)
Government*Inequality	0.002	-0.000	0.009	-0.003
	(0.00)	(0.01)	(0.01)	(0.01)
Constant	0.053***	0.050***	0.063***	0.043***
	(0.01)	(0.01)	(0.01)	(0.00)
Village fixed effects	Yes	Yes	Yes	Yes
Random effects at the experimental session	Yes	Yes	Yes	Yes
Observations	2880	960	960	960
$\chi 2$	378.229	116.512	301.275	192.896
p-value	0.000	0.000	0.000	0.000

Note: Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.13. Multilevel mixed models of the effect of treatments on the Gini coefficient, calculated without sanctioning costs.

	(1) Total sample	(2) Brazilian site	(3) Peruvian site	(4) Indonesian site
Monitoring	-0.003**	0.001	-0.006**	-0.004***
	(0.00)	(0.00)	(0.00)	(0.00)
Community	-0.012***	-0.006**	-0.021***	-0.010***
	(0.00)	(0.00)	(0.00)	(0.00)
Government	-0.013***	-0.011***	-0.016***	-0.012***
	(0.00)	(0.00)	(0.00)	(0.00)
Inequality	0.002	0.004	0.004	-0.002
	(0.00)	(0.00)	(0.00)	(0.00)
Monitoring*Inequality	-0.001	-0.002	0.000	-0.001
	(0.00)	(0.00)	(0.00)	(0.00)
Community*Inequality	-0.002	-0.004	0.004	-0.005
	(0.00)	(0.00)	(0.00)	(0.00)
Government*Inequality	0.000	-0.001	0.004	-0.003
	(0.00)	(0.00)	(0.01)	(0.01)
Constant	0.051***	0.048***	0.048***	0.043***
	(0.00)	(0.00)	(0.00)	(0.00)
Village fixed effects	Yes	Yes	Yes	Yes
Random effects at the experimental session	Yes	Yes	Yes	Yes
Observations	2880	960	960	960
$\chi 2$	567.439	139.496	453.102	173.948
p-value	0.000	0.000	0.000	0.000

Note: Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

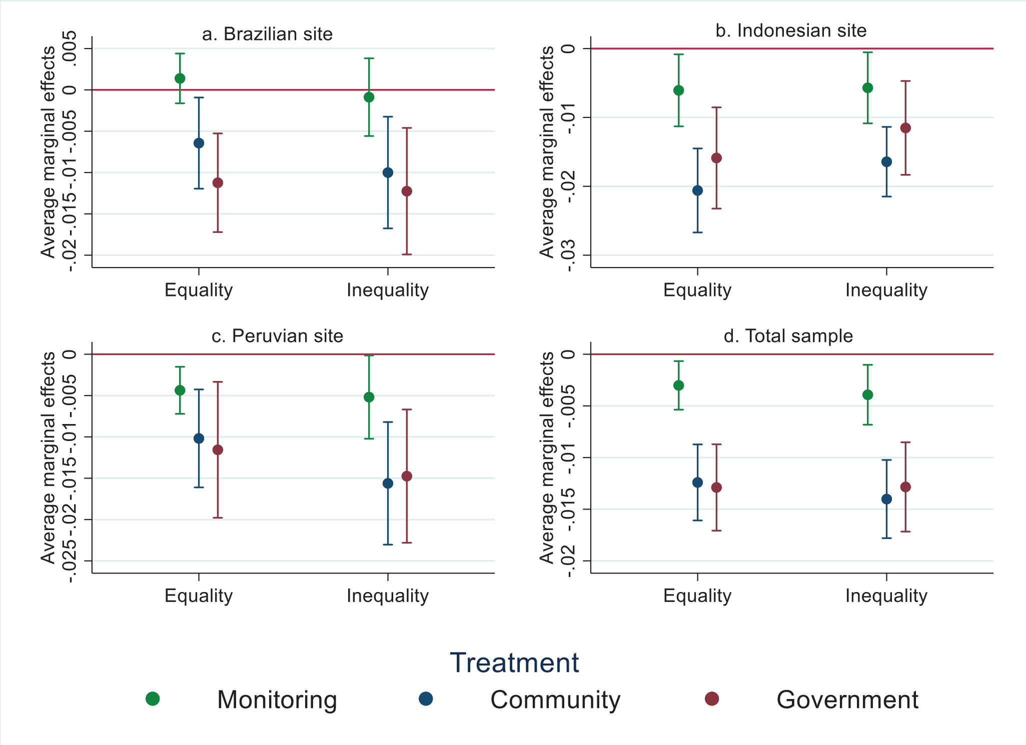


Figure 2.6. Average marginal treatment effects of Public monitoring, Community and Government enforcement on the Gini coefficient of earnings without punishment costs, for Equal and Unequal groups and by country. Vertical lines represent 95% confidence intervals of the coefficients.

Table 2.14 presents the results of the analysis of fairness responses. Participants from both Indonesia and Peru were more likely to select that both Community and Government enforcement are equally fair, most likely due to their community level management of forest resources. This suggests that the perception of fairness is a pre-experimental feature. However, if a participant was sanctioned by Government during the experiment, he/she was less likely to choose Government as the fairer institution. Furthermore, a higher impact of Community enforcement on the participant (i.e., a higher decrease in deforestation levels during the Community enforcement as compared to the baseline stage) implied a higher probability of selecting Community enforcement as fairer. Finally, if Community enforcement reduced inequality compared to a baseline stage, Government enforcement was less likely to be chosen as fairer. Thus, the probability of choosing Government enforcement was not only dependent on the performance of Government enforcement itself, but also on the performance of the Community enforcement.

Table 2.14. Multinomial logit model of perceived fairness. The base category is "Both Government and Community enforcement are equally fair".

	(1) "Community is fairer"	(2) "Government is fairer"
Average individual deforestation	0.021 (0.094)	0.008 (0.083)
Peer punishment received (#)	0.022 (0.049)	-0.013 (0.040)
Government sanctions received (#)	-0.072 (0.092)	-0.229*** (0.085)
Reduced inequality in community stage (dummy)	-0.433 (0.354)	-0.501* (0.295)
Reduced inequality in government stage (dummy)	0.287 (0.352)	-0.023 (0.305)
Age	-0.003 (0.008)	-0.015* (0.008)
Gender	-0.063 (0.243)	-0.269 (0.241)
Reduced deforestation in government stage compared to baseline	-0.200 (0.130)	0.027 (0.126)
Reduced deforestation in community stage compared to baseline	0.255* (0.143)	-0.031 (0.133)
Player with low deforestation capacity (dummy=1)	-0.407 (0.321)	-0.019 (0.282)
Player with high deforestation capacity (dummy=1)	0.057 (0.318)	0.293 (0.289)
Indonesia	-2.229*** (0.476)	-1.159*** (0.425)
Peru	-2.623*** (0.397)	-2.511*** (0.365)
Order enforcement (1=Government first)	0.008 (0.276)	-0.172 (0.238)
Constant	2.434*** (0.745)	3.951*** (0.703)
Observations	698	
$\chi 2$	153.999	
p-value	0.000	

Note: Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A5. Probit and Tobit Models

Table 2.15 presents the regression analysis of Table 1.1 in the main text, conducted using ordered probit models rather than linear models. The results of both types of models are consistent. Table 2.16 presents the regression analysis of absolute and relative deforestation by subsamples based on deforestation capacity. Heterogeneous treatment effects can be analyzed by examining either interaction effects or sub-samples; both are presented as robustness check. Consistent with the results presented in the main text, Government sanctioning had the strongest effect for all participants, but it was strongest in participants with high deforestation capacity, followed by participants with medium and low deforestation capacity. Monitoring had a positive effect on cooperation levels of the participants with both medium and high deforestation capacity, but no significant effect on participants with low deforestation capacity.

Table 2.15. Multilevel mixed effects ordered probit with random effects at the individual and experimental session level, for the total sample and per country.

	(1) Total sample	(2) Brazilian site	(3) Indonesian site	(4) Peruvian site
Monitoring	-0.18^{***}	-0.00	-0.40^{***}	-0.16^{**}
	(0.04)	(0.05)	(0.07)	(0.07)
Community	-0.76^{***}	-0.56^{***}	-1.08^{***}	-0.71^{***}
	(0.06)	(0.08)	(0.10)	(0.12)
Government	-1.14^{***}	-1.08^{***}	-1.07^{***}	-1.30^{***}
	(0.06)	(0.10)	(0.10)	(0.10)
Inequality	0.09	-0.01	0.40^{***}	-0.13
	(0.11)	(0.10)	(0.14)	(0.22)
Village fixed effects	Yes	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes	Yes
Observations	17280	5760	5760	5760
p-value	0.00	0.00	0.00	0.00

Note: Clustered standard errors at the experimental session level in parentheses * p < 0.1, ** p < 0.05, *** p < 0.01

Table 2.16. Multilevel mixed effects ordered probit (columns 1 to 3) and left censored tobit (columns 4 to 6) models of absolute and relative deforestation, with random effects at the individual and experimental session level, per type of participant.

	Absolute deforestation			Relative deforestation		
	(1) Low deforestation capacity	(2) Medium deforestation capacity	(3) High deforestation capacity	(4) Low deforestation capacity	(5) Medium deforestation capacity	(6) High deforestation capacity
Treatments						
Monitoring	-0.080	-0.193***	-0.218***	-0.026	-0.052***	-0.048***
	(0.069)	(0.058)	(0.068)	(0.018)	(0.016)	(0.016)
Community	-0.637***	-0.722***	-0.958***	-0.191***	-0.194***	-0.217***
	(0.095)	(0.097)	(0.094)	(0.027)	(0.027)	(0.022)
Government	-0.990***	-1.204***	-1.228***	-0.293***	-0.327***	-0.276***
	(0.104)	(0.085)	(0.111)	(0.031)	(0.027)	(0.026)
Trust variables						
Group trust	-0.110	-0.067*	0.079	-0.031*	-0.017	0.020
	(0.067)	(0.039)	(0.074)	(0.017)	(0.011)	(0.017)
Village trust	0.108	-0.039	-0.034	0.032	-0.009	-0.009
	(0.067)	(0.045)	(0.093)	(0.019)	(0.012)	(0.021)
Risk aversion						
Extreme or severe	-0.105	-0.274*	0.126	-0.068	-0.077*	0.035
	(0.256)	(0.148)	(0.211)	(0.069)	(0.042)	(0.047)
Intermediate or moderate	-0.223	0.098	-0.182	-0.070	0.022	-0.037
	(0.194)	(0.121)	(0.213)	(0.058)	(0.032)	(0.047)
Social preferences (0=Egalitarian)						
Altruist	0.107	-0.009	-0.348	0.039	0.001	-0.086*
	(0.160)	(0.132)	(0.216)	(0.045)	(0.036)	(0.047)
Inconsistent	0.426	-0.119	-0.041	0.113	-0.031	-0.017
	(0.308)	(0.216)	(0.350)	(0.076)	(0.059)	(0.082)

Spiteful	-0.092	0.354**	-0.123	-0.020	0.100**	-0.023
	(0.332)	(0.156)	(0.182)	(0.096)	(0.042)	(0.041)
Individual characteristics						
Age	0.001	0.009**	-0.002	-0.000	0.002**	-0.001
	(0.005)	(0.004)	(0.007)	(0.002)	(0.001)	(0.002)
Gender (1=Male)	-0.100	0.162	-0.281	-0.002	0.040	-0.071*
	(0.194)	(0.120)	(0.174)	(0.050)	(0.032)	(0.039)
Close family (#)	0.780***	0.268	0.528*	0.227***	0.076	0.119
	(0.250)	(0.217)	(0.320)	(0.066)	(0.060)	(0.074)
Distant family (#)	-0.096	0.090*	-0.124	-0.031	0.023*	-0.030
	(0.106)	(0.050)	(0.098)	(0.031)	(0.013)	(0.022)
Friends (#)	0.100*	0.045	-0.089	0.033**	0.012	-0.022
	(0.057)	(0.040)	(0.094)	(0.016)	(0.011)	(0.020)
Intercept				0.430** (0.194)	0.621*** (0.114)	0.800*** (0.126)
Order of enforcement (1=Government first)	✓	✓	✓	✓	✓	✓
Village dummies	✓	✓	✓	✓	✓	✓
Round control (1 to 6)	✓	✓	✓	✓	✓	✓
Experimental session dummy (1 to 5)	✓	✓	✓	✓	✓	✓
Random effects						
Experimental session level	Yes	Yes	Yes	Yes	Yes	Yes
Individual level	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4320	8640	4320	4320	8640	4320
χ2	3034.16	4527.01	4706.61	3309.80	1345.31	8110.46
p-value	0.00	0.00	0.00	0.00	0.00	0.00

Note: Clustered standard errors at the experimental session level in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.0$

B. Supplementary Methods

In this section, additional description of our experimental design and sample is provided. Section B1 presents a detailed explanation of the experimental setting and the payoff functions. Section B2 describes the methods for the elicitation of risk and social preferences. In section B3 the result of the balance test is presented, followed by a detailed description of the study sites in section B4. Section B5 and B6 contain the experimental material and field protocol.

B1. The experimental design and data analysis

The FFE conducted was similar to a Common Pool Resource (CPR) game (Ostrom, 1990; Ostrom, 2006), framed as a public good game with extraction. We recreated a situation where an individual's deforestation reduced the forest stock and thus its public benefits. In the experiment, participants shared access to a common forest which provided collective benefits in the form of a group-based (collective) PES. By introducing collective PES at the baseline stage, we did not evaluate the additionality of the collective PES as compared to an "open access" situation, as the topic has been explored in other experimental studies (Kaczan et al., 2017; Andersson et al., 2018a; Handberg and Angelsen, 2019; Ngoma et al., 2020). At the beginning of the experimental session each participant was assigned a unique plant or animal symbol (see section B6), which had to remain secret throughout the session. In addition, for stages 2, 3 and 4, each participant was assigned a unique letter of the alphabet, known only to the participant himself/herself and the experimenter. Whenever information on individual decisions was disclosed to the group, the information was shared using the participant's secret letter. This procedure maintained anonymity and ensured that any behaviour was triggered by the decision per se and not by any relationship with the decision-maker, or caused any post-experiment conflicts among participants. Participant's letter was changed at the start of each stage in order to minimize spillover effects across stages, for example, assigning punishments in the Community stage based on a player's decisions in previous stages. Before the baseline stage started, the structure and procedures of the game were carefully explained, and any questions raised were addressed (see section B6 for the script).

<u>The payoff function and optimal strategies during enforcement stages</u>

The decision-making process can be formalized as follows. Let x_{it} be the number of plots of forest that the player decides to deforest, and x_{-it} the deforestation of other players. The benefits of deforestation are set to 1 and let δ be the individual earnings from the collective PES obtained from the standing forests (i.e., the marginal per capita return of the public good). With a total stock of forest plots equal to *S*, and given the maximum number of plots to be deforested $\overline{x_i}$, the monetary pay-off during the baseline stage for participant *i* in round *t* is:

$$\pi_{it} = x_{it} + \delta(S - x_{it} - \sum x_{-it}) \;\; ; \;\; x_i \leq \overline{x_i} \tag{2.2}$$

The two conditions necessary for creating a social dilemma are that: (i) the return of deforestation of forest land (x_{it}) is higher than the individual return of the collective PES ($\delta < 1$), and (ii) the individual return from deforestation is lower than the group benefits from the collective PES ($\delta n > 1$), with *n* being the number of forest users. Thus, the parameters must satisfy the condition $\delta < 1 < n\delta$. The levels of the parameters were set at $S = 60$, and $\delta = 0.4$. Considering individual pay-off maximizing users, the Nash Equilibrium, defined as the set of strategies where no one has an incentive to change their behaviour, occurs when everyone maximizes deforestation. However, from the perspective of the group, the best strategy is when there is no deforestation at all, as it yields higher returns than the Nash equilibrium. Thus, self-maximizing individual strategies lead to outcomes that are not socially optimal and lower individual earnings. The participant had three options: fully cooperate (zero deforestation), no cooperation (deforest the maximum amount allowed), and partially cooperate (deforest less than the maximum amount).

During the government stage, with a fine f being implemented with probability p, the payoff function becomes:

$$\pi_{it} = \begin{cases} x_{it} + \delta\,(S - x_{it} - \sum x_{-it}) - pf x_{it} & \text{if } x_{it} > 0 \\ \delta\,(S - \sum x_{-it}) & \text{if } x_{it} = 0 \end{cases} \tag{2.3}$$

A non-deterrent sanction does not enforce cooperation and is such that $(1 - \delta) > p * f$. A non-deterrent sanction was such that $p = \frac{1}{3}$ and $f = 1.5$. This means that in every round, two out of the six participants were randomly selected to be monitored and deducted 15 points for each unit of land they had chosen to convert to agriculture.
To engage in non-cooperative behaviour (i.e., deforestation), the gains from deforestation must not be lower than the cost of being punished by the peers. Deforestation of x_{it} forest plots generates a gain of $(1 - \delta)\, x_{it}$ compared to conservation (cooperative behaviour). During the Government enforcement stage, partial cooperation is never an optimum for a pay-off maximizing participant, given that the marginal benefit of deforestation is higher than the marginal benefit of conservation. Thus, the player had the alternative of fully contributing or not.

- If full contribution payoff is: $\delta\,(S - \sum x_{-it})$.
- If no contribution, expected payoff is: $x_{max} + \delta\,(S - x_{max} - \sum x_{-it}) - pf x_{max}$

A risk neutral player is indifferent between the two outcomes when $(1 - \delta) = p * f$. A marginal penalty rather than a fixed penalty was chosen so that participants with low, medium and high deforestation capacity face the same treatment (the same incentives to cooperate).
The payoff function during the Community enforcement stage was:

$$\pi_{it} = x_{it} + \delta\,(S - x_{it} - \sum x_{-it}) - k * p_{i,-i} - p_{-i,i} * 3 * k \tag{2.4}$$

k is the cost of assigning punishment points to the peer, and p is the number of punishment points assigned by participant *i* to/from peers. The optimal strategy during this stage depends on the expectations of receiving punishments from other participants. In particular, if there are n' participants punishing the free riders, the player suffers a monetary loss of $n'3k$. Thus, free riding is advantageous as long as $(1 - \delta)\, x_{it} > n'3k$. Assuming risk neutral players, the optimal strategy depends on the expected number of punishers and the punishment ratio:

$$x_{it} = \frac{n'3k}{(1 - \delta)} \tag{2.5}$$

Given the parameters, $(1 - \delta) = 0.6$, thus the highest possible gains of non-cooperation for a participant with low, medium and high deforestation capacities are 2.4 (= 4 x 0.6), 3.6 (= 6 x 0.6), and 4.8 (= 8 x 0.6). Each received punishment reduces earnings by 3, thus the optimal strategy for participants with low deforestation capacity is to fully cooperate if they expect *at least* 1 member to punish him/her for a non-cooperative behaviour, while medium and high capacity participants should fully cooperate if they expect to be punished by *at least* 2 members of the group if they choose to deforest. If a participant expected the probability of being punished to depend on *how much* they deforested, partial cooperation can be observed. Similarly, participants who are more risk averse would opt more for more cooperative choices.

B2. Measuring preferences and fairness perceptions of participants

Before the forest experiment, social preferences were elicited to control for behavioural differences across participants, as social preferences affect cooperation rates in public goods games (Ngoma et al., 2020). Risk aversion preferences were also elicited as these influence the behaviour towards probabilistic punishment. These were measured by presenting participants with six lottery choices (Binswanger, 1981).

Social preferences measure an individual's inclination to values such as equality, altruism, or reciprocity (Fehr and Schmidt, 1999). Each participant received three binary choice sets regarding prosocial, envy and sharing preferences (Fehr et al., 2013). The social preferences were measured following Fehr et al. (2013). Before the CPR game, participants were presented with three choice sets: a prosocial game, an envy game, and a sharing game. The choices sets are presented in Table 2.*17*, in terms of points earned.

Table 2.17. Lottery choice sets used to elicit risk preferences.

Preference elicitation	**Choices**	**Own payoff**	**Other's payoff (points)**	**Outcome**
Prosocial Game	A	200	200	If choose A, prosocial preferences
	B	200	0	
Envy Game	C	200	200	If chooses D, envious preferences
	D	200	400	
Sharing Game	E	200	200	If chooses E, sharing preferences
	F	400	0	

Based on these three choice sets, participants were classified into three categories following Fehr et al. (2013): Egalitarian, Spiteful, Altruist. A fourth category called "Inconsistent" comprised the participants who did not classify into the first three categories. To avoid random answers, one randomly chosen choice set counted towards the participant's final payoff.

To control for risk preferences, participants were presented with six lottery choices, following (Binswanger, 1981). This risk eliciting method relies on making a single choice among a set of alternatives with dichotomous "good luck" or "bad luck" gambles. This method is more easily understood by subjects who might struggle with descriptions involving varying probabilities and payoffs such as in Holt and Laury (2002). Simpler measures produce coarser categorization, but decisions are substantially less noisy (Dave et al., 2010). The choice set presented to the participants is shown in Table 2.18, and the probabilities were adjusted to fit the ones set during the Government enforcement stage. The values were calibrated such that there is a switching point for the dominant strategy when participants have a modestly high relative risk aversion. Participants were classified into three main categories, analyzing risk preferences as a dummy variable: Extreme or Severe, Intermediate or Moderate risk aversion, and Slight to Neutral risk aversion.

Table 2.18. Lottery choice sets used to elicit risk preferences.

Risk aversion preference	"Bad luck" outcome. (p=1/3)	"Good luck" outcome (p=2/3)
Extreme	0	0
Severe	-20	100
Intermediate	-40	260
Moderate	-80	400
Slight to neutral	-160	480
Neutral to preferring	-240	520

To measure trust, we distinguished between general village trust and experimental group level trust (Sturgis and Smith, 2010). The first question stated: "*In general, do you trust people in the group that you participated with?*" and the second was "*In general, do you trust people in your community?".* Participants were able to give an answer from 1 to 5 depending on the proportion of people they trusted. The family ties and social relationships amongst participants in the same experimental session were recorded and included as covariates, including the number of close and distant family members, and whether there were any close friends within the same group. Close family members were defined as the immediate family (including grandparents or grandchildren) and close friends as a trusted individual with whom there is regular social interactions and is not a family member. Participants were also asked about their perceptions of fairness of Government as compared to Community enforcement. These questions related to their experience during the game, but participants unavoidably also brought their real-life experience into the responses. The question was: "*Which rules, "Community Enforcement" or "Government Enforcement" do you consider to be fairer, or do you think they are the same?*".

B3. Balance tests

Table 2.19 indicates that the distribution of covariates is balanced across treatments except for the risk and social preferences, which is included as control in the subsequent analyses. The model was not jointly significant (χ^2=37.4, p=0.67).

Table 2.19. Balance tests across experimental session types. The experimental sessions with Equal capacity to deforest and Community enforcement first are the base category.

	(1) Equal group, Government first	(2) Unequal group, Community first	(3) Unequal group, Government first
Age	-0.005 (0.009)	0.002 (0.008)	-0.006 (0.008)
Gender (Male=1)	0.263 (0.231)	0.218 (0.227)	-0.026 (0.246)
Risk preferences			
Extreme or severe	-0.045 (0.275)	-0.156 (0.302)	-0.237 (0.290)
Intermediate or moderate	-0.493* (0.277)	-0.198 (0.235)	-0.273 (0.279)
Social preferences			
Altruist	-0.661*** (0.238)	-0.159 (0.213)	-0.620*** (0.217)
Egalitarian	-0.390 (0.383)	-0.341 (0.348)	-0.548 (0.389)
Spiteful	-0.329 (0.346)	-0.180 (0.380)	-0.298 (0.332)
Trust			
Village trust	0.077 (0.091)	0.095 (0.087)	0.042 (0.106)
Group trust	0.131 (0.083)	0.017 (0.084)	0.093 (0.077)
Family and friendships within the group			
Number of close family members	-0.005 (0.458)	0.469 (0.381)	-0.416 (0.929)
Number of distant family members	-0.014 (0.143)	-0.072 (0.143)	0.083 (0.149)
Number of close friends	0.025 (0.074)	-0.038 (0.074)	-0.027 (0.090)
Country			
Indonesia	0.193 (0.664)	0.105 (0.635)	0.030 (0.657)
Peru	0.174 (0.678)	0.117 (0.715)	-0.106 (0.683)
Intercept	-0.233 (0.741)	-0.281 (0.651)	0.323 (0.743)
Observations	17 280		
χ2	37.429		
p-value	0.672		

Note: Clustered standard errors by experimental session in parentheses, * p < 0.1, ** p < 0.05, *** p < 0.01.

B4. Study sites and sample selection and description.

To participate in the experiment we prioritized smallholders who had participated in the household surveys conducted by CIFOR's Global Comparative Study, which had randomly selected a sample of village households and gathered socioeconomic data and village level data from June 2018 to January 2019 (Sills et al., 2017). A total of 120 experimental sessions took place, 40 in each country (5 sessions in each of the 8 villages). At least half of the villages in each country have received or are in the process of receiving monetary incentives for forest conservation, while also facing regulations and enforcement mechanisms.

Field researchers held a village meeting one day before the experimental sessions were run, or the morning before the first experimental session. During the meeting, the field researcher provided a list of all the households that had participated in the household survey of the Center for International Forestry Research (CIFOR) in 2018 and asked whether they were present and willing to participate in the experiment. From the list of available households willing to participate, a random sample was selected to participate in the experiment, until a sample of 30 participants per village was reached. Once the 30 participating households were identified, each household was assigned a random number from 1 to 5 to define in which experimental session they would participate. Field researchers avoided having two close family members in the same experimental session. For ethical considerations, before presenting the experiment in the village meeting, field researchers asked the village leader whether there were households among those participating in the survey that have conflicts or disputes. If there were, these households did not participate in the same experimental session, to avoid post-experimental conflicts. There was no criteria for selecting which household head (husband or wife) went into which experimental session, as it was randomly picked. If the randomly picked household member was unavailable, then an alternative household member was asked to participate, with the criteria that the household member had to be above 16 years of age. If in a village it was hard to reach the sample of 30 different households, then it was allowed to select a second adult household member (e.g., the spouse) from one of the households already participating. The households that could have two participants in different experimental sessions were chosen randomly. No experimental session was conducted unless there were exactly 6 participants. If on the day and time of the experiment there was one participant missing, it was possible to include another villager. If possible, field researchers were asked to include one from the original sample of the 2018 survey, but if this was difficult to find, then it was possible to include any other randomly picked villager.

Table 2.20 presents the basic socioeconomic characteristics and preferences at the study sites. The average age of the participants was 44, and 52% of them were men. Agriculture in Peru is conducted by smallholders in parcels smaller than 2 ha, and the most important commercial crops are plantain and papaya. Agriculture in Indonesia is conducted by smallholders in plots less than 1 ha and up to more than 10 ha, with average plot size being ca. 2 ha per household. Crop income comes mostly from cacao, cassava, and corn. In Peru and Indonesia, livestock is comprised of smaller animals (chickens, ducks and pigs). In Brazil, a small portion of total land area is dedicated to crops (ca. 3%), and cattle ranching is the main economic activity in the study site. The largest income share in the Indonesian site is wages, mainly from working at nearby palm oil plantations located outside the village boundaries, while in Peru the most important source of income is from fishing. This wide range of socioeconomic and forest management contexts are captured and controlled for by including village fixed effects in all the models.

Regarding social and risk preferences, most of the participants (46.7%) showed egalitarian preferences, while 29% had altruistic preferences, 13.1% were spiteful and 11.2% were inconsistent. The risk preferences were more evenly distributed across participants, with 31.5% of participants being extreme or severely risk averse, 34.2% being intermediate or moderately risk averse, and 34.4% being slightly risk averse or neutral. Two important differences stand out. First, in Peru, there is a higher proportion of risk neutral individuals (42.1%) than in Indonesia (32.5%) or Brazil (28.3%). Second, Brazil exhibits a higher proportion of altruistic individuals (33.7%) compared to Indonesia (29.6%) and Peru (23.7%). Finally, trust levels, either at the community or the experimental session, were generally higher in Brazil as compared to Peru and Indonesia.

Table 2.20. Key characteristics of the participants in the three study sites. Country averages are presented with standard errors in parenthesis. Anova test were conducted for continuous variables and Pearson's chi-squared for discrete variables.

Variable	Brazilian site	Indonesian site	Peruvian site	Total sample	p-value
Income shares (% of total income)					
Forest income	4.9 (0.1)	6.7 (0.2)	12.1 (0.1)	9.4 (0.2)	<0.001
Fishing	0.13 (0.01)	9.0 (0.2)	30.4 (0.3)	12.7 (0.1)	<0.001
Crops	16.2 (0.1)	9.7 (0.2)	20.3 (0.3)	15.3 (0.2)	<0.001
Livestock	47.4 (0.4)	4.7 (0.2)	6.4 (0.1)	17.6 (0.2)	<0.001
Wages	8.5 (0.2)	45.6 (0.5)	19.5 (0.3)	25.7 (0.2)	<0.001
Other	22.8 (0.3)	24.3 (0.3)	11.2 (0.2)	6.8 (0.1)	<0.001
Land use					
Forest land used or owned (ha)	44.8 (0.5)	0.63 (0.02)	1.1 (0.03)	13.4 (0.2)	<0.001
Agricultural land in use (ha)	38.7 (0.5)	1.9 (0.05)	1.5 (0.02)	12.2 (0.2)	<0.001
Access to a common forest	No	Yes	Yes	-	-
Household deforestation (ha yr^{-1})	1.8 (0.06)	0.04 (0.01)	0.43 (0.01)	0.68 (0.02)	<0.001
Socioeconomic characteristics					
Age (years)	41.3	43.4	44.2	43.2	<0.001
Gender (1= male)	0.54	0.51	0.50	0.52	<0.001
Social preferences (% individuals)					
Egalitarian	46.2	45.0	48.7	46.7	<0.001
Altruistic	33.7	29.6	23.7	29.0	
Spiteful	12.1	12.5	14.6	13.1	
Inconsistent	7.9	12.9	12.9	11.2	
Risk averse preferences (% individuals)					
Extreme or severe	36.3	32.5	25.8	31.5	<0.001
Intermediate or moderate	35.4	35	32.1	34.2	
Slight or neutral	28.3	32.5	42.1	34.3	
Gini coefficient					
Income	0.53	0.49	0.40		
Assets	0.51	0.54	0.42		
Land assets	0.37	0.65	0.59		
Community trust (% individuals)					
I trust very few or none (<20%)	7.9	11.2	25.8	15.0	<0.001
I trust a few	21.7	18.7	38.3	26.2	
I trust some (~50%)	30.8	30.8	12.5	24.7	
I trust many of them	8.3	17.1	8.7	11.5	
I trust most of them or all (>80%)	31.2	22.1	14.6	22.6	

Experimental session trust (% individuals)					
I trust very few or none (<20%)	8.3	38.7	21.7	22.9	<0.001
I trust a few	17.9	24.2	26.7	22.9	
I trust some (~50%)	21.2	9.6	12.9	14.6	
I trust many of them	7.1	4.2	15.8	9.1	
I trust most of them or all (>80%)	45.4	22.9	22.9	30.4	

B5. Field Protocol and script

In this section, the segments that are contained in brackets and in italics are instructions for the moderators. The rest are the instructions that were read out loud to the participants, in the language of the country (Portuguese, Bahasa Indonesian and Spanish). The payoff of the game was presented in terms of "points", and the exchange rate was set such that the expected average payment would be equivalent to the country's rural daily wage. In Indonesia and Peru, payments were in cash, while in Brazil payments were in-kind, with commonly used commodities, due to security concerns and recent robberies to field researchers in the region. The experimental design followed CIFOR's Research Ethics Review of the process, and before starting the experiment verbal consent was requested from the participants.

Starting the experiment

[The experiment should be set up and prepared before the participants arrive to the experimental session. As participants arrive to the experimental session, fill in and confirm/rectify the "Registration form" with their information. When all the participants have arrived, ask them to draw randomly one of the six figures used to identify participants during the activity (print the document "Figures"). Once they all have a figure, start reading the following script.]

Welcome everyone and thank you for being here today. My name is *[name of moderator].* I and my team *[introduce your team]* are here today as researchers interested in understanding how you make decisions about the land that you manage, so we will be conducting today an activity that allows us to do that. We are not affiliated with the government or any local implementing organization. We will use the information we get today for scientific purposes only, and your name and what you earn in the exercise will not be shared with anyone.

This activity is different from activities that other members of the community have participated in. If you have heard comments from other members of the community, they do not necessarily apply to the activity we will play today.

The experiment will take approximately three hours, and at the end of the activity you will be able to earn some cash. You will earn money, a minimum of *[indicate minimum earnings].* You can consider this as a compensation for your time here. We kindly ask you to remain until the end of the game. If you have to leave before the activity ends, you might do so, but we won't be able to give you a payment.

During the activity, you will be earning "points". Each of you will earn points depending upon your own choices and the choices made by the other members of your group. At the end of the activity these points will be converted to cash at the following rate: 200 points = *[indicate the amount in local currency].* Your earnings in cash will depend on your decisions during the activity but will be approximately equivalent to day's wage *[indicate the amount in local currency].* When we have finished playing the activity, we will conduct a short survey asking you questions about yourself and your experience during the activity, and then you will receive your payment in a sealed envelope. Only you will know how much you earned.

Now, here are general rules that you always have to follow *[refer to the poster with the written rules].* First, we ask you not to talk to each other, and be silent during the activity. We also ask you to keep your decisions anonymous and secret. It is important that nobody knows about the decisions and choices you are making during the activity. If you have any questions during the activity please raise your hand and we will answer them.

Before we proceed, we would like to ask for your consent and agreement to participate in the activity. Do you all agree to participate in the activity? *[Wait for verbal consent].*

Great! Let's begin. From now on, please don't speak to your neighbours and listen carefully. We will distribute one by one the decision cards in which you will have to write your answers. We will explain how to use them as we go along.

Social preferences elicitation

The first thing we will ask you to do is to choose between possible ways to allocate points between yourselves. This is the first opportunity to earn points. In these three questions, you will have to choose how to allocate points between you and another person in this group. Out of those three questions, only one will count for your final payoff. We will select which one counts after you have answered all three questions. *[Distribute decision card.]*

[Explanation of decision sheet with example for visual aid.] Look at your first decision card. In this card, you have to choose ONE of the two options, option A or option B: The option A is that you earn 200 points for yourself and another person from the group, you don't know who, will also earn 200 points. The second option, option B, is that you earn 200 points for yourself and we do not give anything to anyone else in your group.

Raise your hand if you have any questions and we will answer your question [*Leave time for questions and answer.]*

Which option do you prefer? We will give you some time to make your decisions. Mark it with a cross. Please remember that you cannot talk between yourselves. It is very important that the decision is personal and secret. *[Leave time for participants to write their answer.]*

Fold the paper so that no one else sees it and we will come and collect it. *[Wait for every participant to answer and fold their sheet, then collect.]*

[Important: Each time you collect the decision sheets, shuffle them a little bit before handed over to the person entering the data into the PC/tablet, so that the participants see that we don't monitor their individual decisions. The moderator in the PC/tablet has to make sure the decision cards are correctly filled. If they are not, they have to be returned to the participant for corrections.]

[Explanation of decision sheet, with example sheet for visual aid.] On the second decision card, you have to choose between two options, option C or option D. Option C is that you earn 200 points for yourself and another person from the group, you don't know who, will also earn 200 points. The second option, option D, is that you earn 200 points for yourself and another person from your group, you don't know who, will earn 400 points.

Raise your hand if you have any questions and we will answer your question. *[Leave time for questions and answer.]*

Which option do you prefer? We will give you some time to make your decisions. Mark it with a cross. Please remember that you cannot talk between yourselves. *[Leave time for participants to write their answer.]*

Fold the paper so that no one else sees it and we will come and collect it *[Wait for every participant to answer and fold their sheet, then collect.]*

[Explanation of decision sheet, with example sheet for visual aid.] On the third decision card, you have to choose between two options, option E or option F. Option E is that you earn 200 points for yourself and another person from the group, you don't know who, will also earn 200 points. The second option, option F, is that you earn 400 points and we do not give anything to anyone else in your group.

Raise your hand if you have any questions and we will answer your question. *[Leave time for questions and answer.]*

Which option do you prefer? We will give you some time to make your decisions. Mark it with a cross. *[Leave time for participants to write their answer.]*

Fold the paper so that no one else sees it and we will come and collect it. *[Wait for every participant to answer and fold their sheet, then collect].*
Now we will decide which one of these three question will count towards your final payoff. If the dice rolls a 1 or a 2, we will count the decision you took between A and B. If the dice rolls a 3 or a 4, we will count the decision you took between C and D. And if the dice rolls a 5 or 6, we will count the decision you took between E and F *[roll the dice and write the outcome on the excel sheet.]*
Risk preference elicitation
Now we will ask you a fourth question. In this question you can either earn points or lose points. In the decision set, you must choose your preferred option between six lotteries. Each one is offering you two different amounts of points that you can either win or lose. You lose points if the dice rolls a 1 or a 2. You can earn points if the dice rolls a 3, a 4, a 5, or a 6.
In option A, there is no difference in the outcomes, you will not gain nor lose points, no matter what the dice rolls.
In option B, you can lose 20 points if the dice rolls a 1 or a 2, or you can win 100 points if the dice rolls a 3, or a 4, or a 5, or a 6.
In option C, you can lose 40 points if the dice rolls a 1 or a 2, or you can win 260 points if the dice rolls a 3, or a 4, or a 5, or a 6.
In option D, you can lose 80 points if the dice rolls a 1 or a 2, or you can win 400 points if the dice rolls a 3, or a 4, or a 5, or a 6.
In option E, you can lose 160 points if the dice rolls a 1 or a 2, or you can win 480 points if the dice rolls a 3, or a 4, or a 5, or a 6.
Finally, in option F, you can lose 240 points if the dice rolls a 1 or a 2, or you can win 520 points if the dice rolls a 3, or a 4, or a 5, or a 6.
In the decision card, you must write what is your preferred option. You can only pick one. Please raise your hand if you have any questions. *[Leave time for questions and answers.]*
Please now write down what option do you prefer. *[Leave time.]* We will now collect your decision cards and roll the dice. *[Collect them and roll the dice to have the final number of pots.]*

I. BASELINE STAGE
Now, we will start with the main activity about forest management. Please listen carefully. While we explain, please raise your hand if you have any questions or if you don't understand something we are saying.
Let's imagine this is a forest *[point to the board].* There is a conservation project whose objective is to stop forest cover loss by stopping the expansion of agricultural areas. For this, the project is asking you to not create any agricultural plots in this area. As a compensation for conserving the forest, the organization will give to the group a payment for every plot you conserve as forest. Each square on this board represents one plot *[point to the "forest" board]* that can be converted to agriculture. Each plot has a size of 0.5 hectares.
You can think of this payment as a compensation for the benefits that the forest has on our climate, for example in regulating the rainfall or in capturing and storing carbon. For each plot that is conserved as forest, thanks to this project the group will earn 24 points, and the points will be divided between all of you equally, which means each of you will earn 4 points.
During the activity, you will have to individually and secretly make a choice about how many plots of agriculture you would like to have in this forest. For each agricultural plot you decide to have, you will earn 10 points, for yourself. You can think of these points as the value of the crops that you grow in the land that you converted, such as plantain, rice, or from the pasture for the cattle

[*read only the option(s) that is/are relevant for the site. Do not focus on a single crop, allow the participant to think about all the crops he/she cultivates].* The more land you have in agricultural use, the higher your earnings, but the payment the group receives for the forest is lost, so you will be reducing the earnings of other people in the group.
The earnings from the group will be higher the more forest the group decides to keep. From the perspective of your own benefits you may want to convert as much forest as possible to agriculture, but from the perspective of the benefits that the group will receive you should convert as little as possible.
The activity will consist of 24 rounds, in each round, you will choose how many agricultural plots you want to have. You can think of every round as a year or as an agricultural season. In every round, you have to consider that the agricultural plots that you had the previous year (the previous round) do not longer exist.
Now, we will individually show you one card that assigns to each of you a letter, and that also indicates what is the maximum number of agricultural plots you can have. You will need to memorize this information. *[Go individually to each participant with their respective participant cards – their participant card that has the same figure than their decision cards – show them their letter and the maximum amount of agricultural plots that they are allowed to have. Ask them to memorize this letter or write it down, with their maximum number of plots and not share it with anyone].*
[Read EITHER the E or U instructions depending on the type of experimental session.]

E: EQUAL GROUP TREATMENT
Now, does everyone know which letter they have? If someone does not, please raise your hand. *[Make sure everyone knows their letter.]*
Each of you will make the decision about how many plots of forest to convert to agriculture secretly and individually. You will write down with numbers or bars how many agricultural plots you want to have in the forest *[use the example card as visual aid to indicate how to fill the decision cards.]* Each one of you can have between 0 and 6 plots. Think of this limit of 6 plots as determined by the amount of money and tools you have to clear forest. You can choose any number you want between 0 and 6, but can never exceed your maximum amount of 6.
Once all of you have made your decisions, using your letters we will write and show on the board *[point to the board]* what was the total number of agricultural plots in the group, and the average number of agricultural plots.
In the box on the right hand side of the decision card, we also ask you to write down how many agricultural plots you think people will have, on average *[use the example card as a visual aid on where to fill the decision card]*. How many agricultural plots do you think or expect any other person of this group to have? It has to be a number between 0 and 6. Do you think they will have on average, 0 agricultural plots, or 1, or 2, or 3, or 4, or 5, or 6? This information we will not share with the group, but the two persons in this group who make the best guesses during this activity will earn 300 extra points.
We have handed you a sheet of paper *[indicate payoff table]* with a table that indicates how many points you will earn depending on your decisions and others' decisions *[indicate the payoff table].* In this table, you can see how many points you receive at the end of each round. The points you earn depend on two things: the number of agricultural plots you individually decide to have, and the number of agricultural plots that the others decide to have. The numbers that are inside the table, in white, correspond to the points that you can earn in each round.

The columns in this table indicate your decision on how many agricultural plots you want to have. We can see that as the number you choose to have of agricultural plots increases, the higher are the points that you receive *[indicate by moving to the right along the columns of the table].*
The rows indicate the choice of others. As you can see, as the total number of agricultural plots in the group increases, the points you receive decreases *[indicate by moving down along the rows of the table].* Here, you can also see the average extraction that there is in the group *[indicate the blue column].* Let's now walk through two examples of how to use the table.
For the first example, let's imagine that you decide to have 3 agricultural plots. If others in the group also have 3 agricultural plots, your payoff is 246 *[point to the payoff table].* If however, people decide to have a total of 23 agricultural plots, your payoff is reduced to 166 *[point in payoff table].*
For the second example, let's imagine that the others in the group decided to have 16 agricultural plots *[point to the payoff table].* If you choose to not convert any forest, your payoff is then 176. But, if you choose to convert your maximum number, with 6 agricultural plots, your payoff increases and is 212 *[point to the payoff table].*
Now, can you tell us what would be the number of points you receive if you have 4 agricultural plots and the others have in total 21 agricultural plots? Please indicate that number in your payoff table. If you have a question, please raise your hand, we will come and explain again *[Wait for the participants to give an answer, and assist and check the answer individually. Correct answer is 180. Go individually to each participant and make sure they understood how the table works. If they didn't explain again.]*

U. UNEQUAL GROUP TREATMENT

There are two types of participants the group. Three of you have randomly received a card that allows them to convert between 0 and 4 plots from forest to agriculture, and three of you have randomly received a card that allows to convert between 0 and 8 plots from forest to agriculture. Think of this difference as the differences in the money and tools people have to clear forest. Some of you have more money and better tools to clear forests, some of you less.
Now, does everyone know which letter they have and the maximum number of agricultural plots they can have? If someone does not, please raise your hand. *[Make sure everyone knows their letter and maximum capacity.]*
Each of you will make the decision about how many plots of agriculture you want to have secretly and individually, and you will indicate it in this decision card *[use decision card as a visual aid on how to fill the decision cards].* You will write down with numbers or bars how many plots of agriculture you want to have in that forest. Some of you can write any number between 0 and 4, while others can write any number between 0 and 8. You can choose any number you want but can never exceed your maximum amount of either 4 or 8 plots of forest. If you write 6 when in reality you can only convert up to 4 plots of forest, then we will take that as a 4.
Once all of you have made your decisions, using your letters we will write and show on the board *[point to the board]* what was the total number of agricultural plots in the group, and the average number of agricultural plots.
In the box on the right hand side of the decision card, we also ask you to write down how much forest do you think that other people will change to agriculture, on average *[use the example card as a visual aid on where to fill the decision card].* How many agricultural plots do you think or expect any other person of this group to have? It has to be a number between 0 and 6. Do you think they will have on average, 0 agricultural plots, or 1, or 2, or 3, or 4, or 5, or 6? This information we

will not share with the group, but the two persons in this group who make the best guesses during this activity will earn 300 extra points.
We have handed you a sheet of paper *[indicate the payoff table]* with a table that indicates how many points you will earn depending on your decisions and others' decisions *[indicate the payoff table sheet].* In this table, you can see how many points you receive at the end of each round. The points you earn depend on two things: the number of agricultural plots you individually decide to have, and the number of agricultural plots that the others decide to have. The numbers that are inside the table, in white, correspond to the points that you can earn in each round.
The columns in this table indicate your decision on how many agricultural plots you want to have. We can see that as the number you choose to have of agricultural plots increases, the higher are the points that you receive *[indicate by moving to the right along the columns of the table].*
The rows indicate the choice of others. As you can see, as the total number of agricultural plots in the group increases, the points you receive decreases *[indicate by moving down along the rows of the table].* Here, you can also see the average extraction that there is in the group *[indicate the blue column].* Let's now walk through two examples of how to use the table.
For the first example, let's imagine that you decide to convert forest to have 3 agricultural plots. If others in the group also have 3 agricultural plots in total, your payoff is 246 *[point to the payoff table].* If however, people decide to convert a total of 23 plots of forest, your payoff is reduced to 166 *[point in payoff table].*
For the second example, let's imagine that the others in the group decided to convert forest to have16 agricultural plots *[point to the payoff table].* If you choose to not convert any forest, your payoff is 176. If, however, you choose to have your of 8 plots of agriculture your payoff increases to 224, and if you can have a maximum of 4 plots your payoff increases to 200 *[point to the payoff table].*
Now, can you tell us what would be the number of points you receive if you have 4 agricultural plots and the others have in total 21 agricultural plots? Please indicate that number in your payoff table. If you have a question, please raise your hand, we will come and explain again *[Wait for the participants to give an answer, and assist and check the answer individually. Correct answer is 180. Go individually to each participant and make sure they understood how the table works. If they didn't explain again.]*

[Read the following in ALL the sessions]
It is very important that during the activity you do not talk to each other or communicate in any way, because we want to keep your choices secret. Any questions on the rules of the activity? We will soon start. *[Leave time for questions and answer publicly.]*
Let's start!

[Play first 6 baseline rounds. In each round the moderator has to wait for everyone to make their decisions and fold their decision cards before collecting. When all the decision cards are collected, it is <u>very important</u> to supervise that the numbers that are read are correct (the person entering the data into the PC/tablet can also check for that). Also, keep a poker face when entering the data, and don't show and reactions to the choices made. At the end of each round, ONLY the total, and average plots of forest converted to agriculture should be announced to the group.]

II. PUBLIC MONITORING STAGE
For the next six rounds, each of you will make the decision about how many plots of agriculture you want to have secretly and individually, and you will indicate it in the decision card [*use*

decision card as a visual aid on how to fill, and repeat explanation is needed]. The difference now is that once we collect all of your decisions, we will share on this board the information about how many agricultural plots each of you decided to have. We will share this information using your letter, that is why it is important to keep it secret and not share it with anyone. We will also share the total amount of agricultural plots and the average amount of agricultural plots in the group. *[Play second set of 6 rounds. In each round the moderator has to wait for everyone to make their decisions and fold their decision cards before collecting. When all the decision cards are collected, it is <u>very important</u> to 1) shuffle them before computing and sharing the results, to preserve anonymity, and 2) supervise that the numbers that are read are correct (the person entering the data into the PC/tablet can also check for that).*

[For rounds 13 to 24, the field work supervisor will either play first the stage "Community Enforcement" or "Government Enforcement", depending on the treatment assigned to the session].

III. GOVERNMENT ENFORCEMENT

For the next six rounds, we will play with different letters. We will come to you and individually show you one participant card that assigns to each of you a new letter, from G to L. The participant card also indicates what is the maximum number of plots you can convert from forest to agriculture, as a reminder. *[Go individually to each participant with their respective participant cards – their participant card that has the same figure than their decision cards – show them their letter and the maximum amount of agricultural plots that they are allowed to have. Ask them to memorize this letter or write it down, and not share it with anyone. Be careful that no other participant see the card.]*

[Read following sentence only if Community Enforcement was played first] In the next rounds, there is no community regulation anymore, you are not allowed to assign deduction points to each other.

For these rounds, we will add a new rule, we can call it "Government Enforcement". The community/group is still receiving payments from an organization for each plot of forest that you keep, as a compensation for the benefits that the forest provides to all of us. Thus, the points the group receives for the remaining forest are the same (24), as well as the points that you receive if you decide to convert forest to agriculture (10). However, now the government will be monitoring whether you convert forest to agricultural plots and can fine and sanction those who do that. Think of this as the *[relevant government agency],* that wants to stop the loss of forests and help you obtain the highest group benefits from the conservation project.

In your decision card, you will write down how many plots you wish to convert from forest to agriculture, and how much you think the others are converting. We will then share with the group how many plots each participant changed from forest to agriculture, by using your secret letter. We will also share the total number of agricultural plots of the group, and the average number of agricultural plots.

Once we have shared this information, the government will come and monitor who converted forest, in every round. The government does not have the time or resources to monitor all of you all the time, so the government will only monitor two persons. Who gets monitored depends on what the dice *[show the dice]* rolls. We will assign one number to each letter. The letter "G" corresponds to number 1, "H" corresponds to number 2, "I" corresponds to number 3, "J" corresponds to number 4, "K" corresponds to number 5, and "L" corresponds to number 6 *[write on the board which letter corresponds to which number, so that it is clear for everyone].*

We will roll the dice twice. If your participant number comes out as an outcome of the dice, then you will be monitored. If you are monitored, and you converted forest to agricultural use, you will have to pay a fine of 15 points for every plot of forest you decided to convert to agriculture. If you are monitored and did not convert any land from forest to agriculture, you will not be fined.
For example, let's imagine in one round participant H chose to convert four units of land from forest to agriculture. Once we have shared the group's total agricultural plots, and how many agricultural plots each of you have, we will roll the dice for the first time. The dice rolls a 2, which corresponds to the letter "H". This means that the government will monitor participant "H". Because he/she converted four plots of forest, he/she will lose 15*4=60 points of whatever she earned in that round. If the participant 2 is not monitored (if the dice rolled another number), then nothing will be subtracted from his/her account, because they government didn't monitor him/her. A participant can only be monitored and fined once. For example, if the second time we roll the dice it also gives us a "2", then we will roll it one more time until we have a different number than "2", since participant H is already monitored. Let's imagine the second time we roll the dice the outcome is a 6. This corresponds to participant "L". Let's imagine that participant "L" converted 2 plots of forest to agriculture. Thus, he will be fined with a total of 15*2=30 points.
If you are the participant that is monitored and fined, do not show any reaction. If you react, the other participants will know which letter you have and it will no longer be secret. Is all of this understood? *[Give time to questions and answers]*
We will play the next 6 rounds with these rules.
[Play the 8 rounds of "Government Enforcement". Moderator has to wait for everyone to make their decisions and fold their decision cards before collecting. In each round, when all the decision cards are collected, it is <u>very important</u> to shuffle them before computing and sharing the results, to preserve anonymity. At the end of each round, the individual number of agricultural plots of each participant is shared publicly, as well as the total and average number of agricultural plots. Roll the dice so that at least one other participant from the group can see the result, to increase transparency.]

IV. COMMUNITY ENFORCEMENT
For the next six rounds, we will assign you different letters, from the M to the R. We will come to you and individually show you one participant card that assigns to each of you a new, different, letter. The participant card also indicates what is the maximum number of plots you can convert from forest to agriculture, as a reminder. *[Go individually to each participant with their respective participant cards – their participant card that has the same figure than their decision cards – show them their letter and the maximum amount of agricultural plots that they are allowed to have. Ask them to memorize this letter or write it down, and not share it with anyone. Be careful that no other participant see the card.]*
[Read following sentence only if Government Regulation was played first.] In the next rounds, there is no government regulation any longer, so it will not be monitoring your community anymore. The rule we will play now can be called "Community Enforcement". The community/group is still receiving payments from an organization for each plot of forest that you keep, as a compensation for the benefits that the forest provides to the group. Thus, the points the group receives for the remaining forest are the same (24), as well as the points that you receive if you decide to convert forest to agriculture (10). However, now each of you will be have the opportunity to fine the other participants, depending on how much you disapprove of their decisions during the activity. Think of this as members of your community regulating what other members are doing with the forest.

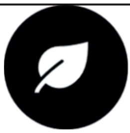
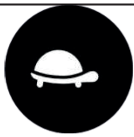
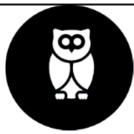
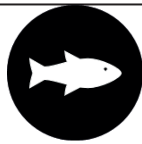
With this rule, each round we play will have two parts. First, as before you will make your decisions about how much you want to convert and what you think the others will convert. After that, we will collect your decision cards. We will then share with the group how many plots each participant converted from forest to agriculture, by using your letters. We will also share the total number of agricultural plots of the group and the average number of agricultural plots. Again, we will only use your letter, so it is important to keep it secret.
Next, you will have the opportunity to fine other participants if you think a participant should be fined. If you want to fine other participants, it will cost you 10 points per fine. The participant that you fine will lose 30 points. You should indicate with a cross to which participant you want to fine in the sheet we are distributing you *[distribute punishment cards and indicate where to place the cross using the example].* If you do not want to fine anyone, do not mark any cross.
For example, if two participants fined you, then you will lose 60 points (30*2). If you also decided to deduct points to another participant, it will cost you 10 points. You can fine a maximum of three participants, and if you do that, it will cost you 30 points. And, you cannot fine yourself! So, we are asking you to write down your letter in the final box of this decision sheet, above the black square. If you do not want to fine anyone, do not mark any cross.
Is this understood? *[Give time to questions and answers.]*
We will play the next 6 rounds with these rules.
[Play the 6 rounds. In each round, the moderator has to wait for everyone to make their decisions and fold their decision cards before collecting. In each round, when all the decision cards are collected, it is <u>very important</u> to shuffle them before computing and sharing the results, to preserve anonymity. At the end of each round, the total and average units of forest converted should be announced to the group as well. Participants will then make their decision on who to fine. Once these decision cards are collected and shuffled, the moderator will publicly display and write on the board which letter decided to assign deduction points to which other participant(s) letter (s). Thus, participants can know which participant assigned punishments to other participants.
It is very important that during this stage the word "punishment" is never said. At the end of this stage, moderator should have collected from each participant eight decision cards and eight "punishment" cards]

V. Post-experiment instructions
We have reached the end of the activity. We want to thank you very much for participating. We hope you enjoyed the activity. This was just an experiment and it does not necessarily relate to how you make decision in real life. We will soon proceed with the payment in cash. Before we do this, we would like to have a short debriefing session and then, you will have to answer a short questionnaire, while we prepare the payments.
[Have a short debriefing session where participants can share what they were thinking when making the decision and answering the activity, and how closely this resembles their reality or not. Do not disclose or say anything about what is the best strategy of the game, what they should have played, etc.]
We kindly ask you again to not speak about the activity we played with the rest of the group members and with people outside this group. This is because we do not want for people who have not played the activity to know the rules of the activity if they play it at some point in the future. We want stress that it is important that you do not reveal the letter to anyone, so that your decision remains private and secret. After we have left, you can talk about the experiment between yourselves, but remember that no one has the right to know how you played in the activity, or how much you earned. The earnings each of you will receive do not differ that much.

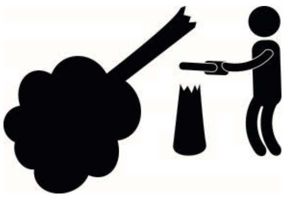
We will soon proceed to payment. While we are making the payment and conducting the individual surveys, all the other participants must remain silent.
[One or two enumerators should help out with the survey, while the coordinator gives out the payments. The surveys have to be conducted with enough distance from the rest of the group so that none of the other participants hear what other respondent is answering. When participant has finished answering the survey, he/she can proceed with the FRS to receive his/her payment].

B6. Experimental material

To play the game, the participants received four types of decision cards: 1) decision cards on how much to deforest, 2) decision cards on how much to "punish" - these were only used in the "Community Enforcement" stage, 3) decision cards on social preferences, and 4) one decision card about risk preferences.
In the decision cards, only the figure of the participant was presented, not the letter. This is to preserve the anonymity of the participants, and to decrease spillover effects across stages. This implies that the matching of the player letter, the figure, and the ID of the participants was only known to the experimenter. Each figure used and corresponding letters are in the figure below.

					
A, K, N	**B, I, P**	**C, H, Q**	**D, L, M**	**E, J, O**	**F, G, R**

Deforestation decision cards. In these cards, the participant used lines or a number to indicate how many units of land he/she wanted to convert from forest to agriculture and what he expected form other members. Participants received 24 cards, one by one, for each round.

 	**My agricultural plots**	**Others' agricultural plots**

Punishment cards. Participants received six decision cards (one for each of the 6 rounds of the stage "Community Enforcement"). This card was used by a participant to inform the moderator to which other player he/she would like to assign punishment points. When a participant wanted to assign deduction points to another participant(s), he/she marked with a cross the letter of the participant(s) he/she wanted to sanction. They could assign deduction points to up to three different participants (i.e., they could not assign deduction points to the same player more than once per round), and could not assign deduction points to themselves.

					
	M	**N**	**O**	**P**	**R**

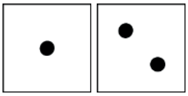
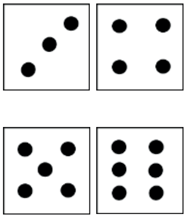
Social preferences cards. To solicit social preferences, participants were presented with the following three sets of choices for allocating points to themselves and to a random partner (indicated by an question mark) in their experimental group. The participant had to mark with a cross which option does he/she prefer.

			**MY CHOICE (One cross)**
A	**200**	**200**	
B	**200**	**0**	

			**MY CHOICE (One cross)**
C	**200**	**200**	
D	**200**	**400**	

			**MY CHOICE (One cross)**
E	**200**	**200**	
F	**400**	**0**	

Risk preference cards. Risk preferences measured an individual's attitude towards risky outcomes. The probability of a bad luck outcome was 1/3, and the probability of good luck outcome is 2/3, to match the probability of being monitored by the Government during the experiment

	**Bad luck (points lost)** 	**Good luck (points gained)** 	**MY CHOICE**
A	0	0	
B	-20	100	
C	-40	260	
D	-80	400	
E	-160	480	
F	-240	520	

3 Paper II

"Quand on a beaucoup médité sur l'homme, par métier ou par vocation, il arrive qu'on éprouve de la nostalgie pour les primates. Ils n'ont pas, eux, d'arrières pensées » Albert Camus, La chute

When one has meditated a lot about humans, by profession or by vocation, it can happen to end up feeling nostalgia for primates. They do not have, them, thoughts at the back of their minds

Will peer punishment protect tropical forests? Multi-country evidence from a framed field experiment.

Julia Naime[1,2] and Arild Angelsen[1,2]

[1] School of Economics and Business, Norwegian University of Life Sciences (NMBU), PO Box 5003, 1432, Ås, Norway
[2] Center for International Forestry Research (CIFOR), Bogor, Indonesia
Emails: *julia.del.carmen.naime.sanchez.henkel@nmbu.no; arild.angelsen@nmbu.no*

Abstract

This paper reports on patters and impacts of peer punishment in a framed field experiment about forest conversion, in situations with either homogeneous or heterogenous agents. The experiment included 720 forest users in Brazil, Indonesia and Peru. Our first research question is to examine the relationship between first order (the appropriation problem of a common pool resource) and second order (peer punishment) cooperation, resulting in the classification of participants into six different behavioural classes. We find that a small share (18.2%) of the participants behaved as self-interested payoff maximisers. The largest sub-group (26.1%) cooperates in both the appropriation and enforcement stages. The remaining participants do not behave consistently across the first order and second order cooperation dilemmas, and we discuss possible motivations. Second, we examine punishment effectiveness: does it change the optimal strategy, and does it lead to reductions in deforestation? We find stark differences across the country sites: in Indonesia, the probability of receiving punishments is roughly twice that in the Brazilian and Peruvian sites. As a result there is no incentive for players to deviate from the group average. Receiving prosocial punishment, defined as the punishment of free riders, effectively reduces deforestation, while receiving antisocial punishment increases deforestation. Overall, the effect of agent heterogeneity on peer punishment is small, while important inter-site variation is observed: inequality increases the frequency of punishments only in Indonesia, and it increases the effectiveness of the punishments only in Brazil.

Keywords: field experiments, peer punishment, common pool resources, deforestation
JEL codes: C73, C93, Q23, Q54

3.1 Introduction

The conservation of common-pool resources (CPRs), such as tropical forests, is a key issue of public economics. CPRs create a collective action dilemma because individuals have an incentive to free ride by overexploiting the resource and reducing its collective benefits. Peer sanctioning overuse may be an effective way to ensure sustainable management of CPRs, as it increases the cost of free riding. This is argued in both local observational studies in the Ostrom tradition (e.g., Ostrom et al., 1992) as well as in the experimental literature in the Fehr and Gächter (2000) tradition. By increasing future cooperation peer punishment creates a collective benefit but also entails individual costs to punishers, thus the peer sanctioning of free riders can be viewed as a second-order collective action problem (Ostrom, 1998; Fehr and Gächter, 2002; Rustagi et al., 2010).

Experimental (mostly lab) studies report on several shortcomings of peer punishments. One is the collective action problem itself: since punishment is costly to the punisher it is also subject to free riding (Ozono et al., 2017). Further, antisocial (i.e., punishment of cooperators) and retaliation punishments exist (Herrmann et al., 2008), and the 'overuse' of punishments may reduce the net benefits as compared to an open access situation (Ostrom et al., 1992; Cason and Gangadharan, 2015). In addition, peer punishment is vulnerable to the establishment of good social norms and monitoring networks. Bad social norms or imperfect monitoring can prevent effective sanctioning and reaching the social optimal outcome (Abbink et al., 2017; Shreedhar et al., 2020).

Most studies on peer punishment are lab experiments with university students (e.g., Herrmann et al., 2008; Gachter and Herrmann, 2009). There is a long-standing debate on to what extent results from such lab experiments generalize to other groups, domains and contexts (Levitt and List, 2007). This is particularly important for peer punishment mechanisms as the patterns of punishment and the norms surrounding its acceptability varies widely across cultures (Henrich et al., 2006; Henrich et al., 2010a; Eriksson et al., 2017). Framed field experiments (FFE) can increase external validity as compared to lab experiments, by framing the cooperation problem to a specific domain, and changing the nature of the participant pool from students to actual resource users (Harrison and List, 2004). Only a few FFEs have been undertaken to study monetary peer punishment in the management of common pools. The exceptions include Lopez et al. (2013) on mollusc harvesting in a coastal community of Colombia, Vollan et al. (2019) on tree harvesting in a woodland savannah area of Namibia, and Kaczan et al. (2017) on a collective Payment for Environmental Services (PES) system in Mexico. In Uruguay, de Melo and Piaggio (2015) evaluated the impact of social (non-monetary) punishment among small scale fishers.

This paper reports on a FFE of tropical forest conversion with 720 participants in Brazilian, Indonesian and Peruvian sites. During the FFE, a group of six local forest users faced a social dilemma, framed as a decision on how many plots to covert to agriculture from a common forest. Conserving the forest gave aggregate benefits to the group in the form of a collective PES scheme, but deforestation gave more agricultural income to the participant than the individual loss of PES income.

We first examine the relationship between first order (FO) cooperation (i.e., the conservation of the common-pool resource) and second order (SO) cooperation (i.e., the peer punishment decision). Second, we examine the effectiveness and impact of punishment by evaluating whether the expected gains from free riding are effectively reduced as compared to the open access situation, as well as whether receiving peer punishment actually leads to lower future deforestation. A novelty of our study is that we examine these questions when there is an equal and unequal distribution of endowments. There is only a handful of laboratory experiments, such as Kingsley (2016), Nockur et al. (2021), and De Geest and Kingsley (2021) that have compared the effect of peer punishment on homogenous and heterogenous groups on a Voluntary Contribution Mechanism (VCM) or CPR game, all in laboratory settings. As such, this paper represents the first multi-country FFE of peer punishment in a CPR game with both homogenous and heterogenous agents.

3.2 Theoretical background

3.2.1 The common pool and peer punishment literature

Individuals with access to a CPR face a conflict between individual and collective benefits. Uncoordinated and self-maximizing behaviour will lead to over-exploitation and eventual depletion of the resource, resulting in the well-known ‘tragedy of the commons’ (Hardin, 1968). However, the tragedy is not unescapable: self-governed communities can successfully manage the commons (Ostrom, 1990). Understanding the capacity of such groups to govern themselves is important because formal, external institutions are not always possible or feasible.

Individuals who engage in monitoring and sanctioning activities create a collective benefit, as punishment potentially reduces exploitation of the common-pool resource, thus increasing future collective benefits. Assuming purely self-maximizing individuals, there should be no peer punishment. However, motivated by social preferences such as fairness (Falk et al., 2005) or by negative emotions such as the feeling of being exploited, individuals often engage in the punishment of free-riders and thus enhance cooperation (Fehr and Gächter, 2000; Fehr and Fischbacher, 2004). In a forest management context, evidence shows that individuals who contribute to the first-order collective action dilemma also contribute more to the second order public good of sanctioning peers (Rustagi et al., 2010).

In repeated games, the punishment of free riding bears two main implications: reducing free riders’ payoff and increasing future cooperation through a disciplining effect. Peer sanctioning can, however, also bring negative effects. Players often engage in antisocial punishment, which occurs when cooperating participants are punished (Herrmann et al., 2008). A framed field experiment among Uruguayan small-scale fishers showed that antisocial punishments can, in fact, lead to more extraction by those that have received antisocial punishments (de Melo and Piaggio, 2015).

Antisocial punishments are more frequent when there is an opportunity to retaliate (Nikiforakis, 2008; Engelmann and Nikiforakis, 2015), suggesting the prevalence of anger or revenge emotions as a motivation for antisocial punishments (Fehr and Gächter, 2002). Furthermore, the social norms underlying punishments can be destructive or welfare reducing, promoting bad norms with a grotesque example being honour killings of rape victims (Abbink et al., 2017). Because of the imperfections of peer punishments, the gains from increases in cooperation are not necessarily offset by the costs of punishment (Ostrom et al., 1992; Nockur et al., 2021).

The type, patterns and levels of punishments are essential determinants of peer punishment effectiveness. Experimental research has extensively examined the effects of peer punishment showing how it depends on the cost of the punishment (Sutter et al., 2010; Chaudhuri, 2011); if sufficiently high, it leads to near full cooperation (Nikiforakis and Normann, 2008). Other factors that enhance the effect of peer punishment include previous communication (Ostrom et al., 1992; Koch et al., 2021) and previous trust and experience (Gelcich et al., 2013; Pfaff et al., 2019). In turn, peer sanctioning might lose effectiveness if there are opportunities to retaliate (Nikiforakis, 2008; Engelmann and Nikiforakis, 2015).

While in general lab experiments show that peer punishment increases cooperation (Chaudhuri, 2011), peer punishment can have no effect on certain participant pools. Peer punishment impact greatly depends on the cultural context and participant pool (Henrich et al., 2006; Eriksson et al., 2017). For instance, Gächter and Herrmann (2011) show that peer punishment does not increase cooperation in a Russian subject pool. Cross-cultural lab experiments highlight the importance of culture and contexts in shaping game outcomes (Henrich et al., 2006; Herrmann et al., 2008; Bruhin et al., 2020; Eriksson et al., 2021). The extent to which the nature and magnitude of cross-cultural differences of lab, 'context-free' experiments persists when cooperation problems are framed and conducted with non-student subject pool remains relatively unexplored. We contribute to filling this gap.

3.2.2 Research questions and hypotheses

Our first research question (RQ1) examines the patterns of punishment (i.e., who is punishing who?). In particular, what is the relationship between FO cooperation and SO cooperation? Examining individual patterns of FO and SO cooperation is important because they determine collective outcomes in forest management contexts (Rustagi et al., 2010). Early studies showed a positive relationship between FO and SO cooperation (Falk et al., 2005; Ones and Putterman, 2007). More recently, Albrecht et al. (2018) examined this question in detail, finding overall two behavioural archetypes: the prosocial type, which is the one who engages in FO cooperation as well as prosocial punishment, and free-rider type, who give less contributions in the first-order dilemma while also being non-punishers. Based on the available evidence, we generate a couple of hypotheses.

> *H1.1:* Prosocial punishment dominates. The probability of receiving punishment tends to be higher the higher the deviation from average extraction (Kingsley, 2016; De Geest and Kingsley, 2021).
>
> *H1.2:* FO cooperators are more active in punishing than FO free riders. A central motivation for peer punishment is to sanction the non-cooperative individuals (Falk et al., 2005). FO cooperators are more likely to perceive uncooperative behaviour as unfair, and thus experience stronger negative emotions towards free riders (Fehr and Gächter, 2002).
>
> *H1.3*: FO free riders engage less in second-order cooperation as they have more selfish oriented motivations.
>
> *H1.4:* Antisocial punishment is more likely to be driven by retaliation behaviour (Nikiforakis, 2008).

The second research question (RQ2) examines punishment effectiveness. First, we examine to what extent does the possibility of being punished change the marginal incentives to deviate from the social norm, defining the social norm as the average group deforestation. We pay attention to how marginal incentives vary by country, as experimental research has indicated strong cultural variation in punishment behaviours and acceptability (Herrmann et al., 2008; Eriksson et al., 2017). We also evaluate the impact of punishment on future cooperation, separately analysing the impact of prosocial vs. antisocial punishment. The experimental literature suggests that introducing punishment opportunities increases overall cooperation (Fehr and Gächter, 2002; Chaudhuri, 2011). Antisocial punishment can, however, reduce cooperation (Herrmann et al., 2008; Gächter and Herrmann, 2011), but not always (Vollan et al., 2019). We thus hypothesize:

> *H2.1:* Participant pools with more exposure to collective institutions will contribute more to second order cooperation.
>
> *H2.2:* While prosocial punishment reduces deforestation, antisocial punishment is likely to increase it.

Our third and final research question (RQ3) concerns how the relationship between FO and SO cooperation (RQ1), and peer punishment effectiveness (RQ2), depends on endowment inequality. This in an important contribution to the experimental literature, as there is limited evidence about the interaction between inequality and punishment in CPRs, and the evidence in VCM games is mixed. Some studies find negative effects on cooperation (Nikiforakis et al., 2012; Kingsley, 2016), others find no effect (Nockur et al., 2021). The differences between the studies can stem from different cost punishment ratio, the nature of agent heterogeneity, as well as the type of public good. For example, Nockur et al. (2021) use a cost-punishment ratio of 1:2, while Nikiforakis et al. (2012)

and Kingsley (2016) use 1:3. Further, Nikiforakis et al. (2012) considers heterogeneity in returns from the public good, while Kingsley (2016) and Nockur et al. (2021) consider endowment heterogeneity. De Geest and Kingsley (2021) is the only study that has evaluated endowment heterogeneity in a CPR context, finding that there is more sanctioning in equal settings as compared to unequal.

3.3 Data and methods

3.3.1 Study sites

The FFE was implemented in 24 villages equally split between three study sites in Pará (Brazil), Central Kalimantan (Indonesia) and Ucayali (Peru) between October 2019 and January 2020. Five experimental sessions were conducted in each village, summing up to 30 participants per village and 240 per site. The average age of the participants was 44 years, and 52% of them were men.

At country level, the eight villages share similar socioeconomic and institutional characteristics, as they were selected to evaluate impacts of conservation projects (Sills et al., 2017). However, there are relevant differences across the countries. Forest is owned communally by indigenous communities in Peru, it is owned by the state in Indonesia, while at the site in Brazil land is owned individually by colonist farmers. In Peru and Indonesia, village households have communal institutions to manage forests and each household controls, on average, an area of ~2 ha for subsistence and/or commercial agriculture. In Brazil, households' control, on average, an area of 44.8 ha of forest and 38.7 ha of agricultural land, mostly pastures.

In Brazil and Peru, land tenure is in most cases considered secure, in the sense that collective (Peruvian site) or individual (Brazilian site) boundaries of properties are legally recognized. In contrast, tenure is considered weak in the Indonesian site as village and households do not have legal recognition of the land they manage and forest access is based on local customary laws, which give individuals land claim when they have invested on that land (e.g., planting, clearing land) (Sills et al., 2014).

Deforestation activities by smallholders serve different economic purposes. In Indonesia, the production is mostly for subsistence consumption, while in Peru, and even more so in Brazil, it is conducted for commercial purposes. In our sample, average household deforestation is much higher in Brazil (1.8 ha yr^{-1}) than in Peru (0.43 ha yr^{-1}) and Indonesia (0.04 ha yr^{-1}). The crop income share is higher in Peru (20.3%) than in Brazil (16.2%) and Indonesia (9.7%). Livestock income is dominant in the Brazilian site (47.4%) while it plays a minor role in the Peruvian (6.4%) and Indonesian (4.7%) sites. The largest income share in the Indonesian site is wages (45.6%), mainly from working at nearby palm oil plantations located outside the village boundaries, while in Peru the most important income source is fishing.

Income inequality is highest in Brazil, but inequality in assets and land is highest in Indonesia (see Appendix B, Table 3.7 for a summary table of the characteristics of the study sites). These wide range of socioeconomic and forest management characteristics are captured and controlled for by including village fixed effects in the regression models.

3.3.2 Experimental set up

The FFE design was in the tradition of CPR games (Ostrom, 1990; Ostrom, 2006), framed as a linear public good game with extraction. In the experiment, six participants shared access to a common forest which provided collective benefits in the form of a group-based PES. In each round, participants chose how many forest plots to convert to agricultural plots, reducing the forest area and thus the group benefits. The framing is similar to studies such as Vollan et al. (2019) or Blanco et al. (2016): participants gain private value from resource units appropriated (i.e., the deforested plots), as well on how many forest plots were left standing once all participants had made their decisions.

The experiment consisted of four stages with six rounds each. To conserve anonymity, each participant was represented by a letter of the alphabet, only known to the participant and the experimenter. The letter was changed in each stage to minimize spillovers across stages (treatments). No communication between participants was allowed in order to reduce the risk of losing anonymity during the experiment. Communication was also prohibited to better capture individual motivations for conservation and sanctioning. Thus, we sought to recreate a non-cooperative environment with no capacity to engage in verbal agreements (Ostrom et al., 1992).

In the first stage, we introduced the baseline with the collective action problem. Let x_{it} be the number of plots of forest that the player decides to deforest, and x_{-it} the deforestation of other players. Setting the benefits of deforestation to 1, δ represents the individual earnings from the collective PES, i.e., the marginal per capita return (MPCR) of the public good. With a total stock of forest plots equal to *S*, the monetary pay-off during the baseline stage for participant *i* in round *t* is:

$$\pi_{it} = x_{it} + \delta\left(S - x_{it} - \sum x_{-it}\right); x_i \leq \overline{x_i} \tag{3.1}$$

The two conditions necessary for creating a social dilemma are that: (i) the return of deforestation of forest land (x_{it}) is higher than the individual return of the collective PES ($\delta < 1$), and (ii) the individual return from deforestation is lower than the group benefits from the collective PES ($\delta n > 1$), with *n* being the number of resource users. Thus, the parameters must satisfy the condition $\delta < 1 < n\delta$. In other words, the Nash Equilibrium occurs when everyone maximizes deforestation within their capacity ($\overline{x_i}$), but from the perspective of the group, the best strategy is no deforestation. We also note that Eq. (3.1) implies that the collective benefit is distributed equally among participants. Everyone received the same benefit from each individual plot of agricultural land.

The levels of the parameters were set at $S = 60$, and $\delta = 0.4$ (Chaudhuri, 2011; Ngoma et al., 2020). The stock of forestland *S* was reset in every round, to avoid effects due to accumulated forest loss. We specified that each plot was equivalent to 0.5 ha. Each plot of agricultural land was worth 10 points, while each plot of forest gave 24 points to the group (4 points to each player). In all sessions, each participant had a payoff table indicating his/her earnings as a function of his/her and others' decisions. Visual support was also provided to explain the collective action dilemma, using a cardboard with 60 green squares. Each square represented a forest plot, and showed the group payoff of 24 points, and the individual payoff of 4 points. Whenever deforestation took place, yellow paper stickers indicating the individual payoff of 10 points replaced the green squares.

Inequality in deforestation capacity was introduced by modifying the maximum number of forest plots that a participant could convert to agricultural land, with a between-session design: half of the experimental sessions had inequality in capacity to deforest, and the other half had equality. We will refer to the two as the unequal and the equal groups. In the *Unequal* groups, three randomly chosen low-capacity participants could deforest a maximum of four plots, and three high-capacity participants could deforest up to eight plots. In the *Equal* groups, all participants had a medium capacity to deforest six plots. Thus, the aggregate deforestation capacity was the same between the Equal and Unequal groups.

The experiment strictly focused on the effects of inequality in deforestation capacity by keeping the marginal benefits of deforestation constant and equal across participants, and the same aggregate deforestation capacity across groups: only the distribution of the individual capacity was different. Thus, the cooperation incentives remain the same for every participant. This inequality was framed in terms of household's differences in capital and labour availability for establishing agricultural plots, and not as differences in opportunity costs of conservation (e.g., unequal income in agriculture).

After the baseline, we sequentially introduced three different treatments: (i) individual monitoring of public deforestation, (ii) external punishment and (iii) peer punishment. Individual monitoring treatment was introduced at the second stage, and we randomized between external and peer punishment in the third and fourth stages. Here, we only use data from the baseline and peer punishment stages.

Overall, the design resulted in four types of experimental sessions, depending on whether there was: (i) equality or inequality in the capacity to deforest, and (ii) on whether peer punishment or external sanction treatments were played at stage 3 or 4. We control in all our analysis for the order in which the punishment treatment were played.

The payoff function during the peer punishment stage was as follows:

$$\pi_{it} = x_{it} + \delta\left(S - x_{it} - \sum x_{-it}\right) - k * g_{it} - r_{it} * 3 * k; x_{it} \leq \overline{x_i} \qquad (3.2)$$

Where k is the cost of assigning punishment points to the peer, and g is the number of punishments given by participant i to peers, and r is the number of punishments received by i from peers at round t. A participant can either fully cooperate (zero deforestation), not cooperate (free ride, maximum deforestation), or partially cooperate (deforest less than maximum). The optimal strategy during this stage depends on the expectations of receiving punishments from other participants. Free riding is advantageous as long as

$(1 - \delta)\ x_{it} > r3k$, with r being the number of punishments received. Given the parameters, if they are not punished the highest possible gains of non-cooperation for a participant with low, medium and high deforestation capacities are 24 (=6 * 4), 36 (= 6* 6), and 48 (6 * 8) respectively. The optimal strategy for risk neutral participants with low deforestation capacity is to fully cooperate if they expect at least 1 member to punish him/her for non-cooperative behaviour (i.e converting a positive amount of forest such that $x_{it} > 0$), while medium and high-capacity participants should fully cooperate if they expect to be punished by at least 2 members of the group if they choose to deforest. If a participant expected the probability of being punished to depend on how much they deforested (e.g., above the group average), partial cooperation can be observed. Risk averse participants would also opt more for more cooperative choices.

The payoff of the game was presented in terms of points, and the exchange rate was set such that the expected average payment would be equivalent to the country's rural daily wage. In Indonesia and Peru, payments were in cash, while in Brazil payments were in-kind, with commonly used commodities, due to security concerns and recent robberies of field researchers in the region. At the end of the game and before making the payment, we conducted a post-experimental questionnaire to ask about punishment motivations in an open-ended question. Multinomial logit regressions indicate that the distribution of trust, social and demographic characteristics are balanced across the four types of experimental sessions, except for the risk and social preferences, which are included as control in subsequent analyses (Appendix B, Table 3.8). The experimental design followed the Center for International Forestry Research (CIFOR) Research Ethics Review, and before starting the experiment verbal consent was requested and given by the participants.

3.3.3 Definitions and data analyses

Two main definitions of prosocial and antisocial punishment behaviour are found in the literature (Cinyabuguma et al., 2006). First, prosocial (antisocial) punishment can be defined as punishment given to a participant who has a deforestation rate above (below) the round average. The second definition considers the punisher's own cooperation level rather than the group average: prosocial (antisocial) punishment occurs when the punished participants have higher (lower) deforestation that the punisher. We opted to use the first definition (average group conversion) to focus on group rather than individual norms, as well as it being the more relevant definition to disentangle the impact on cooperation of receiving each type of punishment.

We defined first-order (FO) cooperators as the participants who converted below the group average during the baseline stage, while FO free riders are those who deforested above the group average. We also identified three main types of punishers: (i) the no punisher, who are those with zero punishments assigned during the stage, (ii) the prosocial punisher, those for whom at least half of the punishments were prosocial, and (iii) the antisocial punishers, those for whom more than half of the punishments were antisocial. This classification is similar to Albrecht et al. (2018). Acknowledging that retaliation is a significant driver of punishment decisions, we further defined a retaliation punishment as the one where participants punish an individual who punished him/her in the immediately previous round (punishing the punishers), irrespectively of whether the punishment was pro or antisocial.

We conduct Mann–Whitney U-test to compare and test for significant differences in punishment frequency between equal and unequal groups, and across country sites. To analyse the relationships between first order and second order cooperation (RQ1), we use multilevel linear and double censored Tobit models, with random effects at the individual and experimental session level. Multilevel models allow to take into account the nested nature (individual and experimental session levels) of the observations, and Tobit models allow to fit censored data (Skrondal and Rabe-Hesketh, 2004).

To analyse how punishment opportunities change the incentives to cooperate (RQ2), we examined the determinants of punishment using Poisson multilevel models (Moffatt, 2015). The key dependent variable was the punishments received, and the key independent variable was the player's deviation from the group average. We included the squared term to allow for any non-linearities. We also included an interaction effect with inequality and countries to evaluate for heterogenous effects.

From the models we calculated the expected number of punishments received as a function of the deviation from group average. Based on the estimated expected punishment, we calculated the marginal gains from increasing deforestation, which allowed to estimate the optimal strategy across countries and contexts. Finally, to evaluate the impact of the punishments received on future cooperation, we regressed the change in deforestation from one round to the next, as a function of past punishments received, using a linear specification.

As controls, we included in all our models social and risk preferences measured following Fehr et al. (2013) and Binswanger (1981), respectively. Preferences were measured before the common-pool resource game. In all models, we also controlled for trust preferences at the group and village level, as well as socioeconomic characteristics such as age and gender. We included village fixed effects, a dummy specifying the round of the stage to accommodate for learning effects, a dummy indicating whether the peer punishment or the external sanction was played first, and a dummy indicating the order of the experimental session within the village (from 1 to 5).

3.4 Results and discussion

To answer our first research question (RQ1) regarding the relationship between FO cooperation and SO cooperation, we first present the descriptive statistics and punishment patterns (section 4.1). We then investigate the determinants of giving punishment (section 4.2). To answer our second research question (RQ2) regarding the effectiveness of peer punishment, we examine the determinants of receiving punishments (section 4.3), before evaluating the impacts on deforestation levels (section 4.4). In each result sub-section, we highlight differences between the equal and unequal groups (RQ3).

3.4.1 First order and second order cooperation

An average of 14.4 punishments were given during the six rounds of the peer punishment session (2.4 per round). Among these, 64.4% were classified as prosocial punishments (i.e., the punished participant had forest conversion *above* the group average in that round), while 35.6% were classified as antisocial punishments (punished those with forest conversion *below* the group average).

There are significant differences in overall punishment levels between equal and unequal groups, and across the three country sites, cf. Figure 3.1. Significantly more punishment is given in the Unequal groups (16.9 per session) than in the Equal groups (11.7 per session) (p=0.06, Mann–Whitney U-test). Second, the frequency of punishment in the Indonesian site (21.5 per session) was about twice as high as in the two other study sites (Peru: 10.5; Brazil: 10.9). Combining these two variables, among the Unequal groups in Indonesia punishments were more than three times as frequent as among Equal groups in Peru (25.7 vs. 8.0).

Across the two types of groups and the three sites, there is considerable variation in the share of prosocial punishments, from 55.7% for the Unequal groups in Peru to 71.1% for the Equal groups in Indonesia. Overall, share of prosocial punishments was slightly (and insignificantly) higher in the Equal sessions (66.5%) than in the Unequal sessions (62.9%), indicating that the effect of an unequal setting is in the overall level rather than in the composition of punishments. The share of prosocial to antisocial punishments of approximately 2:1 is within range of what has been previously reported in experimental games (Gächter and Herrmann, 2011).

Recalling our FO cooperation categorization (first order cooperation occurs when deforestation is below group average in baseline stage) and SO cooperation categorization (no punisher, mostly prosocial punisher, or mostly antisocial punisher), we introduce a typology of six different types of players, shown in Table 3.1.

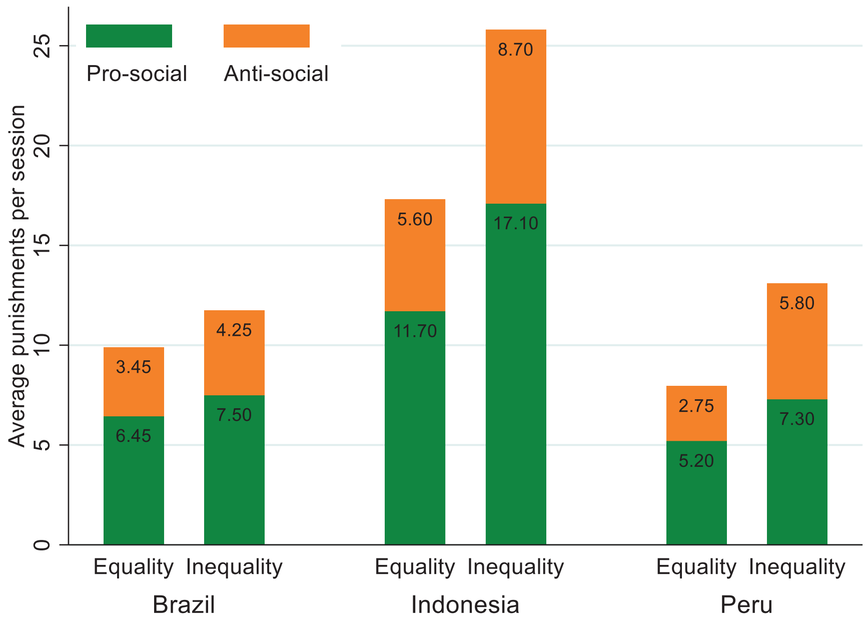


Figure 3.1: Average peer punishments by type of punishment, type of group and country.

Table 3.1.Typology and proportion of players, according to first order (FO) and second order (SO) cooperation. Shares represent average for all groups, while share in parentheses are for equal and unequal groups, respectively.

		Enforcement public good (second order (SO) cooperation)			
		Prosocial punishment (SO cooperators)	**Antisocial punishment**	**No punishment (SO free riders)**	**Total**
Forest public good (first order (FO) cooperation)	**FO cooperator** (equal group, unequal group)	**Homo recriprocans** 26.1% (25.6%, 26.7%)	**Confused** 6.2% (4.4%, 8.1%)	**Benigns** 21.7% (23.6%, 19.7%)	54.0% (53.6%, 54.5%)
	FO free riders (equal group, unequal group)	**Hypocrites** 17.5% (15.0%, 20.0%)	**Saboteurs** 10.3% (10.6%, 10.0%)	**Homo economicus** 18.2% (20.8%, 15.6%)	46.0% (46.4%, 45,6%)
	Total	43.6% (40.6%, 46.7%)	16.5% (15.0%, 18.1%)	39.9% (44.4%, 35.3%)	100%

We label the participants who are both FO and SO cooperators *Homo reciprocans* (Bowles and Gintis, 2002) as they are punishing individuals who are not reciprocating on their cooperative behaviour. They amount slightly more than a quarter of the players (26.1%), and by deforesting less while engaging in prosocial punishments make a double contribution to the group's benefits. Almost as large is the group of the *Benigns* (21.7%), the FO cooperators who did not want to engage in any punishment of their peers. The behaviour of a small group of players (6.2%), the FO cooperators who engaged in punishing antisocially, are labelled the *Confused* as they are not consistent between FO and SO cooperation. Among the FO free riders, a sizeable group are the *Hypocrites* (17.5%); they convert more forest than the group average and punish those that do the same, thus displaying double standards. The *Homo economicus* (18.2%) behave as selfish utility maximisers, free riding on both the FO and SO cooperation games. Finally, the *Saboteurs*, making up about one tenth of the players (10.3%), are FO free riders that also engage in antisocial punishment (punishing low converters), imposing costs on both themselves and the punished.

The study by Albrecht et al. (2018) is relevant to compare these patterns. Similar to their study, we find that the most common types are the Non-punishers (43.6%) and the Prosocial punishers (39.9%). Likewise, the share of FO free riders who are Non-punishers is lower (18.2%) than the share of FO cooperators who are Non-punishers (21.7%). The Antisocial punishers have the lowest share overall (16.5%). Contrary to Albrecht et al. (2018), we find a higher proportion of FO free riders who engage in antisocial punishment behaviour (10.3%) as compared to FO cooperators who do so (6.2%).

The high share of Non-punishers is in contrast with the pattern found with data from the cross-country lab experiment by Herrmann et al. (2008), where the share of non-punishers is 17% (Bruhin et al., 2020). The non-personal setting of lab-experiments, as compared to our field setting, is a plausible explanation of our higher share of non-punishers.

Table 3.1 indicates one major difference between the equal and unequal groups, as already observed in Figure 3.1. Namely, the much higher proportion of individuals who engage in punishments in unequal groups compared to equal groups: 64.8% vs. 55.6%. Interestingly, the difference is not explained by a higher share of *Homo recriprocans* (i.e., the first order and second order cooperators), but rather *Hypocrites* (the FO free-riders who are SO cooperators, meaning they punish prosocially) and *Confused* (the FO cooperators who punish antisocially), while the shares of both *Benigns* and *Homo economicus* are lower in the unequal groups. The higher number of *Confused* and *Hypocrites* in unequal settings is indicative of the ambiguous way in which inequality affects punishment patterns: it increases both the share of prosocial and antisocial punishers.

We also observe substantial differences across the countries for the six typologies (Figure 3.2, see also Table 3.14 in Appendix B). Compared to the two other sites, in Indonesia there is a higher share of *Hypocrite*s (high deforesters punishing fellow high deforesters) and *Homo reciprocans* (low deforesters punishing high deforesters), and a lower share of *Benigns* and *Homo economicus*. This is expected as Figure 3.1 shows that the share of prosocial punishers is higher in Indonesia compared to the Latin American sites. Across the two types of groups, we note for the Indonesian site the much lower share of the

Benigns in the unequal groups, while the shares of both *Homo reciprocans* and *Saboteurs* are higher.

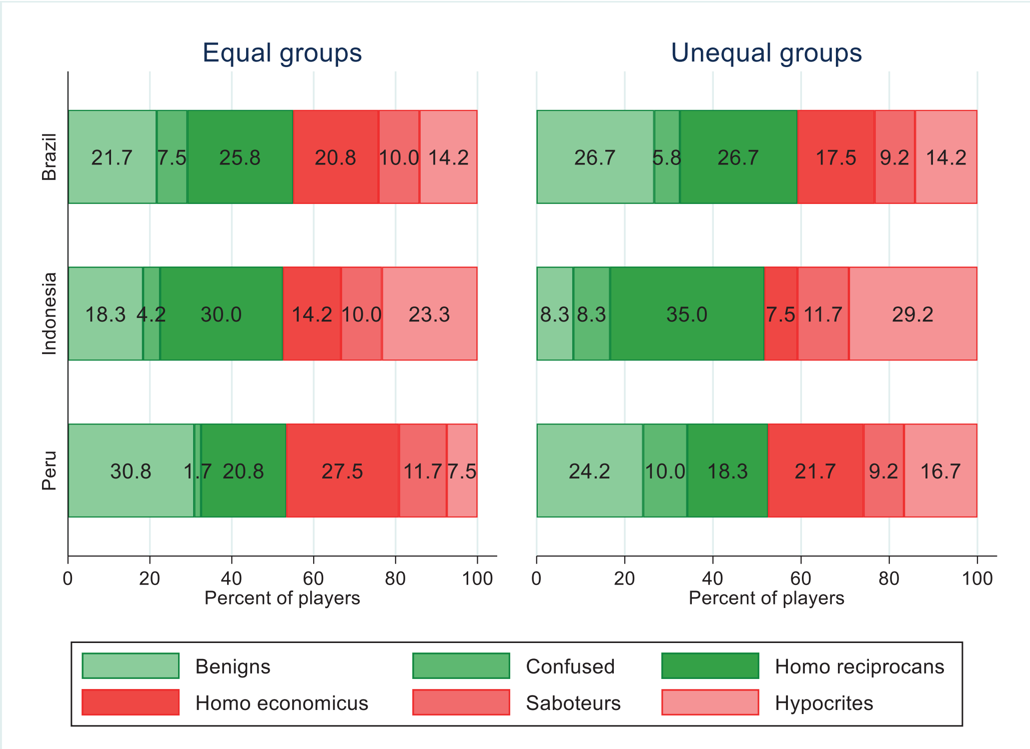


Figure 3.2: Types of players across countries and types of groups.

Homo economicus individuals act according to what is expected from a selfish utility maximiser. The large majority of our sample (81.8%) fails to behave as *Homo economicus*. What are the potential motivations behind each pattern of FO and SO cooperation? Fairness and equity considerations play an important role in explaining cooperation behaviour (Fehr and Schmidt, 1999). The motivation for FO cooperation (*Homo reciprocans', Confused and Benigns'*) can also be driven by a concern of others' payoff or to avoid the guilt of being a free rider (Lopez et al., 2012). Motivations for SO cooperation (*Homo reciprocans* and *Hypocrites*), can include fairness and equity considerations, as by punishing the free riders they reduce their higher (than average) payoff (Falk et al., 2005).

The *Confused* and *Saboteurs*, in turn, who punish the FO cooperators, can be driven by negative emotions such as spite and revenge, and gain utility from reducing other's payoff at a cost to themselves and the others. *Hypocrites* might have a similar motivation to the *Confused* and *Saboteurs,* and driven by spite rather than by fairness concerns, gain utility from reducing the payoff of FO free riders. This motivation is more consistent with their own FO free riding behaviour. Another motivation for antisocial punishment is to target the non-punishers and avoid earning less than non-punishing subjects (Thöni, 2013). A proportion of the punishments can also be linked to random errors by the participants or by spiteful emotions (e.g., Ostrom et al., 1992).

An important motivation in peer punishment is retaliation (Nikiforakis, 2008). However, our total share of retaliation punishments is relatively low (17%, see Table 3.2). Out of the 1 105 prosocial punishments, 155 (14%) can be linked to retaliatory behavior. In turn, out of the 611 antisocial punishments in the sample, 137 (22%) can be associated to retaliatory behavior[13].

Table 3.2. Frequency of punishments, depending on whether they are retaliatory (following punishment in previous round) and prosocial or antisocial. See Appendix B (Tables 3.9. to 3.11) for distribution in each individual country.

	Retaliation	Non-retaliation	Total
Prosocial	155 (9%)	950 (55%)	1 105 (64%)
Antisocial	137 (8%)	474 (28%)	611 (36%)
Total	292 (17%)	1 424 (83%)	1 716 (100%)

Answers from the post experimental questionnaire regarding the main motivations and reasons to punish shows that 59% of the punishers were motivated by prosocial, collective considerations. As stated by some participants: "*some players converted forest land to agricultural land and some other players only converted a few forest plots, which is not fair to other people since the forests were managed together by the community*" and "*there were participants who converted forest land to agricultural land and did not want to share points through PES*".

The second largest share of participants (18%) openly stated that they were being driven by retaliation motivations; they punished because they were punished in previous rounds. A smaller number of participants (8%) mentioned they punished for what could be interpreted as spite, the main reason being "*I just wanted to punish and make others loose points*", or to avoid being punished in the future (5%). A few participants (2%) mentioned punishing because participants were converting too little forest, and thus were perceived as being 'lazy'. Finally, 9% of participants admitted that they punished by error.

3.4.2 Who are the punishers?

We now move to the analysis of who gives punishments. Since the motivations and behavioural patterns are likely to differ between prosocial and antisocial punishments, we conduct a separate analysis for the two punishment types. Tables 3.3 and 3.4 present the results of the tobit regression models. Our main independent variable of interest is the degree of FO cooperation, defined as the deviation from group average in the round. We

[13] We recall that our definition of retaliation is conservative: we defined retaliations as a function of the punishments received in the immediately preceding round. A higher share of antisocial punishment would be included if retaliation included punishments in the previous two rounds.

control for lagged punishment received as retaliation can also be a motivation for punishment.

Table 3.3. Tobit model of giving prosocial punishment. The variable FO cooperator (FO free rider) indicates how much below (above) the group average the player's forest conversion was during the round.

	(1) Total	(2) Brazil	(3) Indonesia	(4) Peru
First order cooperation				
FO cooperator	0.27***	0.13	0.42***	0.30***
	(0.06)	(0.11)	(0.07)	(0.12)
FO free rider	-0.14**	-0.13	-0.13*	-0.19
	(0.06)	(0.11)	(0.08)	(0.13)
Lagged punishments received (1 round)				
Non-retaliatory (#)	0.11**	0.22	0.08	0.19
	(0.06)	(0.21)	(0.06)	(0.12)
Retaliatory (#)	0.08	0.44*	0.02	0.11
	(0.10)	(0.26)	(0.13)	(0.20)
Average group conversion in round	0.28***	0.35***	0.50***	-0.02
	(0.07)	(0.12)	(0.12)	(0.12)
Deforestation capacity				
Low capacity	0.30	-0.37	0.50*	0.34
	(0.19)	(0.40)	(0.27)	(0.29)
High capacity	0.37*	0.37	0.15	0.64**
	(0.19)	(0.34)	(0.23)	(0.32)
Round dummy (1-6)	-0.07**	-0.06	-0.10**	-0.00
	(0.03)	(0.05)	(0.04)	(0.04)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual level covariates	Yes	Yes	Yes	Yes
Observations	3600	1200	1200	1200
Log likelihood	-2137.79	-598.13	-935.81	-547.55
$\chi 2$	193.30	215.72	276.89	223.38
p-value	0.00	0.00	0.00	0.00

Note: Random effects are at the individual and experimental session level. Model is censored at 0 and 3. Clustered standard errors by experimental session in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.4. Tobit model of giving antisocial punishment. The variable FO cooperator (FO free rider) indicates how much below (above) the group average the player's forest conversion was during the round.

Antisocial punishment given	(1) Total	(2) Brazil	(3) Indonesia	(4) Peru
First order cooperation				
FO cooperator	-0.33^{***}	-0.27^{**}	-0.36^{***}	-0.47^{**}
	(0.09)	(0.11)	(0.14)	(0.19)
FO free rider	0.21^{***}	0.14	0.27^{***}	0.12
	(0.06)	(0.11)	(0.09)	(0.13)
Lagged punishments received (1 round)				
Non-retaliatory (#)	0.38^{***}	0.85^{***}	0.19^{**}	0.65^{***}
	(0.07)	(0.15)	(0.07)	(0.19)
Retaliatory (#)	0.42^{***}	0.90^{***}	0.19	0.58^{**}
	(0.13)	(0.22)	(0.17)	(0.26)
Average group conversion in round	0.15^{*}	0.31^{**}	0.46^{***}	0.02
	(0.09)	(0.12)	(0.16)	(0.15)
Deforestation capacity				
Low capacity	0.59^{***}	0.15	0.50	0.71^{**}
	(0.21)	(0.31)	(0.31)	(0.30)
High capacity	0.45^{**}	0.38	0.11	0.63^{*}
	(0.20)	(0.28)	(0.30)	(0.32)
Round dummy (1 to 6)	-0.04	0.06	-0.04	-0.11^{*}
	(0.04)	(0.08)	(0.06)	(0.06)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual level covariates	Yes	Yes	Yes	Yes
Observations	3600	1200	1200	1200
Log likelihood	-1405.42	-389.26	-550.29	-427.23
$\chi 2$	2111.91	535.12	2389.97	2646.13
p-value	0.00	0.00	0.00	0.00

Note: Random effects are at the individual and experimental session level. Model is censored at 0 and 3. Clustered standard errors by experimental session in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A few results emerge from the analysis, related to the hypotheses put forward in section 3.2.2. First, the degree of FO cooperation matters. As expected, participants that deforested less than the group average (FO cooperators) are more likely to give prosocial and less likely to give antisocial punishments. In contrast, FO free riders are more likely to give antisocial and less likely to give prosocial punishments. Further, high average group conversion during the round was associated with more frequent punishments, in particular prosocial ones. This is expected as high group conversion suggests a stronger need for disciplining punishments. The higher the group average, the more likely to be some high converters that are a natural target for prosocial punishments.

While the signs are the same across the three countries, these effects are not significant for all sites. Remarkably, in the Brazilian site, the only significant relationship is that being FO cooperator yields lower propensity to give antisocial punishments. However, in the site with the highest punishment frequency, Indonesia, the four coefficients are slightly larger compared to the other country sites and all significant, suggesting that the fairness and retaliatory motivations for punishing are strongest in this site. As robustness check, we ran supplementary regressions considering the alternative definition of FO cooperation: the dummy variable about first order cooperation during the baseline stage (instead of the deviation from the average deforestation during the round) and find consistent results (see annex, Tables 3.18 and 3.19).

Having received retaliatory punishments has a limited impact on the likelihood of giving prosocial punishments, while the impact of non-retaliatory punishments is significant for the full sample but not the individual country samples. The picture is different for antisocial punishments, where both retaliatory and non-retaliatory punishments seem to induce more antisocial punishments.

Overall our results confirm the hypothesis that the willingness to punish in a forest management context varies across cultures. Eriksson et al. (2017) find that in countries with a collectivistic culture, punishers and non-punishers are rated equally, while in more individualistic cultures punishers are considered *less* favourably. This might explain why in Indonesia participants were less reluctant to engage in punishment behaviour.

The evidence further supports the hypothesis that motivations from antisocial punishment are more likely to be driven and influenced by past punishments received, as compared to prosocial punishment motivations. These results are consistent with previous analysis establishing the relationship between FO and SO cooperation (Albrecht et al., 2018) and retaliatory behaviour and antisocial punishments (Nikiforakis, 2008). The relationship between antisocial punishments and retaliation is stronger than found in other studies, a recent example in Vollan et al. (2019) in Namibia.

Finally, while the existence of inequality increases the number of antisocial and prosocial punishments, whether it is driven by participants with lower or higher capacity to deforest varies across sites (Table 3.3 and Table 3.4). The quantity of prosocial punishments given is higher for high deforestation capacity participants in the total sample and in Peru. In Indonesia however, engaging in prosocial punishment is more likely for low-capacity participants (Table 3.3). Higher antisocial punishment is driven by both low and high deforestation capacity participants and is only significant in the Peruvian site (Table 3.4).

Our results differ from De Geest and Kingsley (2021) where introducing agent heterogeneity in fact reduces the punishment frequency in a CPR game. Possible explanations for these differing results include the nature of the experimental pool (lab with an abstract problem, with undergraduate students vs. a framing with a real problem at hand and with actual forest users) as well as the existence in De Geest and Kingsley (2021) of an 'outside option' for the participants. This outside option represents the income obtained from not appropriating the resource and is different from the collective benefit. De Geest and Kingsley (2021) argue that inequality allowed participants to better coordinate on a contribution norm, while in our case inequality might have had the contrary effect (i.e., hindering coordination on the norm), which would explain why punishments are higher in unequal environments.

3.4.3 What are the gains and costs of free riding?

Figure 3.3 displays how the expected number of punishments received varies by a player's deviation from the group average, distinguishing between the equal or unequal groups (panel a) and the country sites (panel b). The principal conclusion is that high converters are much more likely to be punished by fellow group members. Inequality slightly increases the probability of receiving punishments, but there are no significant differences between the expected punishments of the Equal and Unequal groups (Figure 3 and Table 3.20 in Appendix B). A player that converts four plots more than the group average can expect to be punished by at least two of the five peers. This compares to a likelihood of 0.2-0.3 punishments for those that deforest close to or slightly below the average.

The Brazilian and Peruvian sites manifest the same expected punishment patterns, while there are more expected punishments in Indonesia (Figure 3.3). For Indonesia, 4 punishments are expected if the deviation from the group norm is 4. In contrast, the number of punishments in Brazil and Peru range from 0.27 when there is no deviation from the group norm, to ~1 when there is a deviation from 4 units. Thus, the rate at which the punishment probability increases as a function of deviation from group average in Indonesia is much higher than in Brazil or Peru (Figure 3.3). Indeed, model predictions indicate that the expected number of punishments in Indonesia exceed the experimental limit of maximum 5 punishments received per player.

Our result is consistent with lab evidence that individuals' willingness to punish depends on the intensity of the violation (Fehr and Gächter, 2000; Fehr and Gächter, 2002; Masclet et al., 2003) and in line with the graduated sanctions criteria of successful collective governance (Ostrom, 1990; Wilson et al., 2013). The analysis shows how the targeting of the largest free riders varies by site, being more pronounced in Indonesia. We note that in actual field settings, peer punishment has been shown to have limitations in terms of targeting the largest free riders (Balafoutas et al., 2016).

We also evaluated how expected punishment received varies by the absolute deforestation level instead of the deviation from group average, finding similar results (Appendix B, Figure 3.5): small effect of inequality and significant differences between countries.

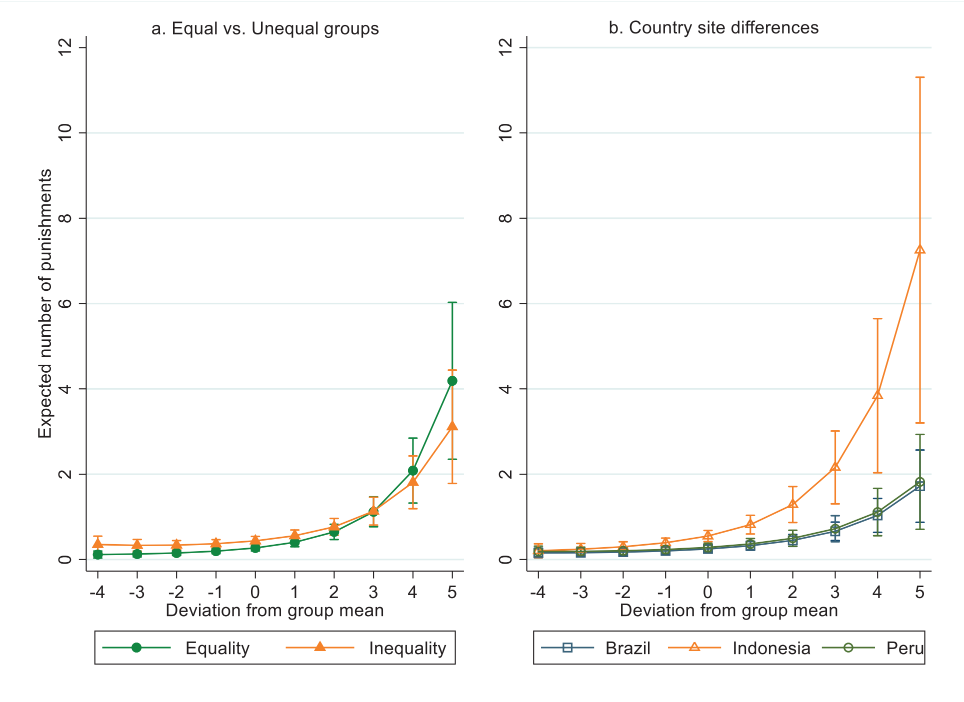


Figure 3.3: Expected punishments received depending on the deviation from the group mean, for equal and unequal groups, and by country. Negative deviations imply antisocial punishments, while punishment of positive deviations implies prosocial punishment. Predictive margins with 95% confidence intervals. See annex (Table 3.20) for the table with full model specification and coefficients.

The expected marginal payoff loss from punishments received (across countries and considering both Equal and Unequal groups) is given in Table 3.5. The numbers indicate the payoff loss from deviating one unit more from the group average. A couple of interesting observations are revealed. First, deforesting less than the group average can lead to a reduction in payoffs, as showed by the significant payoff loss when the deviation from the group average is negative. This is expected as we show in 4.3 that there is a significant amount of antisocial punishment behaviour. Second, the optimal strategy varies across countries. Considering the whole sample, it is optimal for participants to deforested 1 unit more that the group average (the gains of deforesting that unit will be higher than the expected cost). In Brazil and Peru, the optimal deforestation is higher, at 2 units above group average. However, in Indonesia, the optimal strategy is to deforest just at the average group deforestation or unit less, as the marginal loss of deforestation at group average (i.e., 6.27 points) is roughly the same as the marginal gain of 6 points.

Overall, what our results consistently show is that the intolerance to deviation from the social norms (defined by the average group deforestation) is much stronger in the Indonesian site than in the two Latin American sites. A central question is why the marginal benefits of deviating from the social norm are much smaller in Indonesia. Given that in all our models we control for trust, social and risk preferences, the explanation of country differences might reside in a stronger forest enforcement culture in the Indonesian site, which manifests in the stronger reliance in customary rules to manage collective forests in that site (Sills et al., 2014). While tenure regimes (e.g., community vs. individual property) could play a role in the strength of enforcement of a collective agreement, we rule it out as the only explanatory factor, as the Peruvian and Brazilian sites have communal and individual ownership, respectively, yet they have similar tolerance to deviations to the social norm (Table 3.5).

Table 3.5. Marginal loss (in number of points) from punishment, per country and depending on the deviation from the group average. Gray cells indicates that the marginal effect is significantly different from zero. Bold numbers indicate the deviation at which it becomes unprofitable to deviate more from the average, considering that one unit of deforestation brings a marginal benefit of 6 points.

Deviation from group average	Marginal gain from deforestation one more unit	Marginal loss from higher expected punishment (expected increase in punishment * 30 points) from deforesting one more unit			
		Total sample	Brazil	Indonesia	Peru
-4	6	-3.9	-0.51	1.05	-1.74
-3	6	0.15	-0.06	1.59	-0.78
-2	6	0.75	0.33	2.46	-0.033
-1	6	1.56	0.84	3.87	0.69
0	6	2.82	1.59	**6.27**	1.62
1	6	4.98	2.82	10.47	3.09
2	6	**9.03**	5.07	18.09	5.7
3	6	17.13	**9.45**	32.4	**10.83**
4	6	34.35	18.48	60.15	21.69
5	6	73.26	38.1	115.89	46.35

3.4.4 Do punishments make free-riders cooperate?

Table 3.5 demonstrates how the punishment patterns change the 'open access' optimal strategy of deforesting as much as possible. Did the participants act accordingly? Specifically, we ask: how does receiving punishment reduce the forest conversion in the next round?

Figure 3.4 presents the graph on the change in forest conversion, following one or more punishments in the previous round. The pattern in panel (3.4a) is clear; the more prosocial punishments received in a period, the larger the reduction in conversion in the next period. The pattern for the antisocial punishments is different: those receiving 1-2 antisocial punishments do in fact increase the conversion, while those receiving three punishments (only 9 observations) did lower it.

Recalling that a total of 17% (or 292 out of 1 716) of the punishments are retaliatory punishments (Table 3.2), we might expect players to respond differently to them compared to non-retaliatory ones. Panel (b) of Figure 3.4 shows that the disciplining effect of retaliatory punishment is indeed lower than that of non-retaliatory.

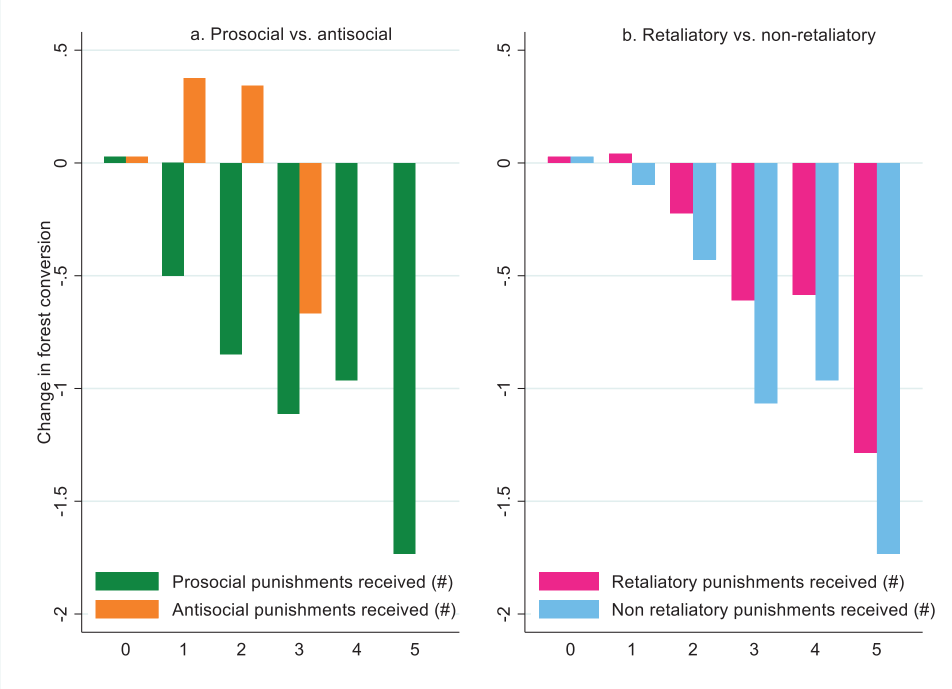


Figure 3.4: Change in forest conversion from previous round, depending on the number of punishments received.

The regression model using the change in deforestation from one round to the next is presented in Table 3.6. Receiving prosocial punishment reduced deforestation by 0.3 units in the next round. In turn, receiving antisocial punishments increased deforestation by 0.1 units. Interestingly, the size of the coefficient of both pro and antisocial was highest in Brazil, even though this is the country site that has no collective land tenure and thus where we could expect lower effectiveness. Another surprising result is that FO free riders reacted more strongly to antisocial punishment. One could thus say that they have a stronger retaliatory behaviour when receiving 'unfair' punishment. We obtain similar results and same conclusions when regressing using the deviation from group average as the dependent variable (see annex, Table 3.*22*).

Thus, in all the country sites, antisocial punishment has two negative effects on the group: besides being costly to both the punisher and the punished, it reduces future cooperation and thus the public good (PES payments). The detrimental effect of antisocial punishment is characteristic of peer punishment contexts in which the decision to punish is uncoordinated (Herrmann et al., 2008; Boyd et al., 2010) and is also observed in the context of social, non-monetary punishments (de Melo and Piaggio, 2015). The results are consistent with the previous lab experiments showing the negative effects of antisocial punishment on cooperation (e.g., Gächter and Herrmann, 2011), but contrast the more recent evaluation of peer punishment in Namibia, where antisocial punishment does not significantly affect cooperation rates (Vollan et al., 2019).

Inequality increased the effectiveness of the punishment, but only in the Brazilian site. This result is consistent with the recent study by De Geest and Kingsley (2021), where they find that inequality increases the effectiveness of peer punishment. However, we show that it cannot be generalized to different subject pools, as the effect was not significant in the Indonesian and Peruvian sites. An additional analysis indicates that the inequality effect in Brazil is dominated, surprisingly, by participants with low deforestation capacity being more responsive to the punishment received (see Appendix B, Table 3.*21*). For these participants, the potential loss from punishment might be perceived as higher even if they deforested their maximum amount of four plots.

Table 3.6. Impact of punishment on deforestation levels.

Dependent variable: Δ Deforestation	(1) Total	(2) Brazil	(3) Indonesia	(4) Peru
Lagged punishments received				
Antisocial (#)	0.119**	0.307*	-0.039	0.163*
	(0.057)	(0.169)	(0.056)	(0.087)
Prosocial (#)	-0.396***	-0.541***	-0.378***	-0.337**
	(0.075)	(0.197)	(0.094)	(0.133)
FO free rider	0.055	0.059	0.136**	-0.022
	(0.034)	(0.086)	(0.054)	(0.035)
Interaction terms				
FO free rider # Anti-social	0.340***	0.407	0.257*	0.470**
	(0.120)	(0.319)	(0.153)	(0.235)
FO free rider # Pro-social	-0.022	-0.083	-0.083	0.126
	(0.083)	(0.257)	(0.104)	(0.128)
Inequality	-0.009	-0.074*	0.037	0.026
	(0.026)	(0.043)	(0.038)	(0.037)
Round dummy (1 to 6)	0.030*	0.030	-0.007	0.065**
	(0.016)	(0.028)	(0.026)	(0.025)
Constant	-0.273***	-0.267*	0.157	-0.479***
	(0.103)	(0.151)	(0.149)	(0.157)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual level covariates	Yes	Yes	Yes	Yes
Observations	3960	1320	1320	1320
Log Likelihood	-6742.343	-2519.961	-1929.124	-2133.711
$\chi2$	208.014	157.291	310.099	255.588
p-value	0.000	0.000	0.000	0.000

Note: Random effects are at the individual and experimental session level. Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A design factor that potentially would increase the effectiveness of peer punishment and reduce antisocial punishment is introducing punishment coordination among participants (Boyd et al., 2010). Various studies have examined variations of the peer punishment rules, by introducing for example, the necessity for individuals to coordinate on punishments decisions by voting on who to punish (Pfattheicher et al., 2018; Nockur et al., 2021) or by introducing communication (Gangadharan et al., 2017). Similarly, delegating the enforcement and punishment decisions to a reduced number of individuals (i.e., leaders or monitors) can solve some of the issues of peer punishment identified in the study, but not always (Nosenzo and Sefton, 2014; Kosfeld and Rustagi, 2015): the effect crucially depends on the motivations of the ones leading the punishment.

3.5 Conclusion

Using a framed field experiment in Brazil, Indonesia and Peru, we extend the literature examining cooperation dilemmas and classify participants depending on how they behave in the first and second order cooperation problems, leading to six distinguishable groups of participants. We found a consistent and positive relationship between first order cooperation (conserving the common-pool resource) and second order cooperation (punishment of free riders). Mirroring that, there is also a consistent relationship between first order free riding and antisocial punishment. Yet a significant proportion of both FO cooperators and FO free riders (40%) do not engage in peer punishment. Our typology of six different player types illustrates, however, the diversity of individual behaviour. Only the *Homo reciprocans* and the *Homo economicus* show consistent behaviour in the two cooperation problems, although the behaviour of the four other groups can be explained by invoking other behavioural motivations.

We further showed that peer punishment can deliver on conservation outcomes and reduce deforestation in the context of collective PES; a large share of participants are willing to punish individuals who are not cooperating. The punished free riders do reduce their deforestation. However, self-enforcement entails a risk of engaging in antisocial behaviour which – besides being costly in itself - has a negative effect on future cooperation. Approximately one third of the punishments were antisocial. Future examination of how patterns of antisocial and prosocial punishment evolve over time can help increase the understanding of peer punishment dynamics.

We finally highlight the important differences in punishment behaviour between the Brazilian, Indonesian and Peruvian sites. Across the sites, Indonesia shows the strongest sanctioning of free riders. Relatedly, the effect of inequality in endowments on peer punishment could not be generalized across sites: it increased the frequency of punishments in Indonesia and Peru, and increased punishment effectiveness only in Brazil. Further research is needed to evaluate whether the impact of inequality is systematically linked to different institutional (i.e., land tenure) as well as cultural context (e.g., fairness considerations).

Acknowledgements

This research is part of CIFOR's Global Comparative Study on REDD+ (www.cifor.org/gcs). The funding partners that have supported this research include the Norwegian Agency for Development Cooperation (Norad, Grant No. QZA-16/0110 & No. QZA-21/0124), International Climate Initiative (IKI) of the German Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU, Grant No. 20_III_108), and the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA) with financial support from the CGIAR Fund Donors. We thank Frode Alfnes for valuable comments to an earlier version of our manuscript. All participants in the study sites in Brazil, Indonesia and Peru are gratefully acknowledged.

Appendix B

Table 3.7. Main statistics and characteristics of each country site.

Variable	Brazilian site	Indonesian site	Peruvian site	Total sample	p-value
Income shares (% of total income)					
Forest income	4.9 (0.1)	6.7 (0.2)	12.1 (0.1)	9.4 (0.2)	<0.001
Fishing	0.13 (0.01)	9.0 (0.2)	30.4 (0.3)	12.7 (0.1)	<0.001
Crops	16.2 (0.1)	9.7 (0.2)	20.3 (0.3)	15.3 (0.2)	<0.001
Livestock	47.4 (0.4)	4.7 (0.2)	6.4 (0.1)	17.6 (0.2)	<0.001
Wages	8.5 (0.2)	45.6 (0.5)	19.5 (0.3)	25.7 (0.2)	<0.001
Other	22.8 (0.3)	24.3 (0.3)	11.2 (0.2)	6.8 (0.1)	<0.001
Land use					
Forest land used or owned (ha)	44.8 (0.5)	0.63 (0.02)	1.1 (0.03)	13.4 (0.2)	<0.001
Agricultural land in use (ha)	38.7 (0.5)	1.9 (0.05)	1.5 (0.02)	12.2 (0.2)	<0.001
Access to a common forest	No	Yes	Yes	-	-
Household deforestation (ha yr^{-1})	1.8 (0.06)	0.04 (0.01)	0.43 (0.01)	0.68 (0.02)	<0.001
Socioeconomic characteristics					
Age (years)	41.3	43.4	44.2	43.2	<0.001
Gender (1= male)	0.54	0.51	0.50	0.52	<0.001
Social preferences (% individuals)					
Egalitarian	46.2	45.0	48.7	46.7	<0.001
Altruistic	33.7	29.6	23.7	29.0	
Spiteful	12.1	12.5	14.6	13.1	
Inconsistent	7.9	12.9	12.9	11.2	
Risk averse preferences (% individuals)					
Extreme or severe	36.3	32.5	25.8	31.5	<0.001
Intermediate or moderate	35.4	35	32.1	34.2	
Slight or neutral	28.3	32.5	42.1	34.3	
Gini coefficient					
Income	0.53	0.49	0.40		
Assets	0.51	0.54	0.42		
Land assets	0.37	0.65	0.59		
Community trust (% individuals)					
I trust very few or none (<20%)	7.9	11.2	25.8	15.0	<0.001
I trust a few	21.7	18.7	38.3	26.2	

I trust some (~50%)	30.8	30.8	12.5	24.7	
I trust many of them	8.3	17.1	8.7	11.5	
I trust most of them or all (>80%)	31.2	22.1	14.6	22.6	
Experimental session trust (% individuals)					
I trust very few or none (<20%)	8.3	38.7	21.7	22.9	<0.001
I trust a few	17.9	24.2	26.7	22.9	
I trust some (~50%)	21.2	9.6	12.9	14.6	
I trust many of them	7.1	4.2	15.8	9.1	
I trust most of them or all (>80%)	45.4	22.9	22.9	30.4	

Table 3.8. Balance tests across experimental session types (multinomial logit specification). The experimental sessions with Equal capacity to deforest and Community enforcement are the base category.

	(1) Equal group, Government first	(2) Unequal group, Community first	(3) Unequal group, Government first
Age	-0.005 (0.009)	0.002 (0.008)	-0.006 (0.008)
Gender (Male=1)	0.263 (0.231)	0.218 (0.227)	-0.026 (0.246)
Risk preferences			
Extreme or severe	-0.045 (0.275)	-0.156 (0.302)	-0.237 (0.290)
Intermediate or moderate	-0.493* (0.277)	-0.198 (0.235)	-0.273 (0.279)
Social preferences			
Altruist	-0.661*** (0.238)	-0.159 (0.213)	-0.620*** (0.217)
Egalitarian	-0.390 (0.383)	-0.341 (0.348)	-0.548 (0.389)
Spiteful	-0.329 (0.346)	-0.180 (0.380)	-0.298 (0.332)
Trust			
Village trust	0.077 (0.091)	0.095 (0.087)	0.042 (0.106)
Group trust	0.131 (0.083)	0.017 (0.084)	0.093 (0.077)
Family and friendships within the group			
Number of close family members	-0.005 (0.458)	0.469 (0.381)	-0.416 (0.929)
Number of distant family members	-0.014 (0.143)	-0.072 (0.143)	0.083 (0.149)
Number of close friends	0.025 (0.074)	-0.038 (0.074)	-0.027 (0.090)
Country			
Indonesia	0.193 (0.664)	0.105 (0.635)	0.030 (0.657)
Peru	0.174 (0.678)	0.117 (0.715)	-0.106 (0.683)
Intercept	-0.233 (0.741)	-0.281 (0.651)	0.323 (0.743)
Observations	17 280		
$\chi 2$	37.429		
p-value	0.672		

Note: Clustered standard errors by experimental session in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.9: Two-sided Mann–Whitney U-tests between the punishment decisions of the Equal and Unequal sessions, separating between prosocial and antisocial punishments (N=120). Standard errors in parenthesis.

Average punishment per round	**Equal sessions**	**Unequal sessions**	**p-value of Mann–Whitney U-test (N=120)**
All punishment	11.7 (1.4)	16.92 (1.9)	0.06
Prosocial punishments	7.78 (0.9)	10.65 (1.2)	0.11
Antisocial punishments	3.92 (0.6)	6.27 (0.9)	0.18

Table 3.10. Brazil - Two-sided Mann–Whitney U-tests between the punishment decisions of the Equal and Unequal sessions. Standard deviations in parenthesis.

Average punishment per round (Brazil)	**Equal sessions**	**Unequal sessions**	**p-value of Mann–Whitney U-test (N=120)**
All punishments	9.9 (1.5)	11.8 (2.9)	0.78
Prosocial punishments	6.7 (1.1)	7.7 (1.8)	0.79
Antisocial punishments	3.2 (0.6)	4.1 (1.2)	0.89

Table 3.11 Indonesia - Two-sided Mann–Whitney U-tests between the punishment decisions of the Equal and Unequal sessions.

Average punishment per round (Indonesia)	**Equal sessions**	**Unequal sessions**	**p-value of Mann–Whitney U-test (N=120)**
All punishment	17.3 (3.3)	25.8 (3.0)	0.03
Prosocial punishments	12.3 (2.3)	17.6 (1.8)	0.07
Antisocial punishments	5.0 (1.3)	8.1 (1.8)	0.29

Table 3.12. Peru - Two-sided Mann–Whitney U-tests between the punishment decisions of the Equal and Unequal sessions. Standard deviations in parenthesis.

Average punishment per round (Peru)	Equal sessions	Unequal sessions	p-value of Mann–Whitney U-test (N=120)
All punishment	8 (1.7)	13.1 (1.1)	0.37
Prosocial punishments	5.2 (1.0)	8.2 (2.1)	0.32
Antisocial punishments	2.8 (0.7)	4.8 (1.1)	0.51

Table 3.13. Kruskal Wallis test to compare punishment across country site. Standard deviations in parenthesis.

Average punishment per round	Brazilian site	Indonesian site	Peruvian site	Kruskal-Wallis test
All punishments	10.9 (1.6)	21.6 (2.3)	10.5 (1.8)	<0.001
Prosocial punishments	7.2 (1.0)	15.0 (1.5)	6.7 (1.2)	<0.001
Antisocial punishments	3.7 (0.7)	6.6 (1.1)	3.8 (0.7)	0.49

Table 3.14. Typology and proportion of players, according to first order (FO) and second order (SO) cooperation, per country and depending on Equal or Unequal context.

First order coop	**FO Cooperators**			**FO Free riders**		
Second order coop	**No punishment**	**Antisocial**	**Pro-social**	**No punishment**	**Antisocial**	**Pro-social**
Type	**Benigns**	**Confused**	**Homo reciprocans**	**Homo economics**	**Saboteurs**	**Hypocrites**
Total	156 (21.7%)	45 (6.2%)	188 (26.1%)	131 (18.2%)	74 (10.3%)	126 (17.5%)
-equal	85 (23.6%)	16 (4.4%)	92 (25.6%)	75 (20.8%)	38 (10.6%)	54 (15.0%)
-unequal	71 (19.7%)	29 (8.1%)	96 (26.7%)	56 (15.6%)	36 (10.0%)	72 (20.0%)
Brazil	58 (24.2%)	16 (6.7%)	63 (26.2%)	46 (19.2%)	23 (9.6%)	34 (14.2%)
-equal	26 (21.7%)	9 (7.5%)	31 (25.8%)	25 (20.8%)	12 (10.0%)	17 (14.2%)
-unequal	32 (26.7%)	7 (5.8%)	32 (26.7%)	21 (17.5%)	11 (9.2%)	17 (14.2%)
Indonesia	32 (13.3%)	15 (6.2%)	78 (32.5%)	26 (10.8%)	26 (10.8%)	63 (26.5%)
-equal	22 (18.3%)	5 (4.2%)	36 (30.0%)	17 (14.2%)	12 (10.0%)	28 (23.3%)
-unequal	10 (8.3%)	10 (8.3%)	42 (35.0%)	9 (7.5%)	14 (11.7%)	35 (19.2%)
Peru	66 (27.5%)	14 (5.8%)	47 (19.6%)	59 (24.6%)	25 (10.4%)	29 (12.1%)
-equal	37 (30.8%)	2 (1.7%)	25 (20.8%)	33 (27.5%)	14 (11.7%)	9 (7.5%)
-unequal	29 (24.2%)	12 (10.0%)	22 (18.3%)	26 (21.7%)	11 (9.2%)	20 (16.7%)

Table 3.15. Frequency of punishments in Brazil, depending on whether they are retaliatory (following punishment in previous round) and prosocial or antisocial. Overall percentage in parenthesis.

	Retaliation	Non-retaliation	Total
Prosocial	35 (8%)	244 (56%)	279 (64%)
Antisocial	39 (9%)	115 (27%)	154 (36%)
Total	74 (17%)	359 (83%)	433 (1005)

Table 3.16. Frequency of punishments in Indonesia, depending on whether they are retaliatory (following punishment in previous round) and prosocial or antisocial. Overall percentage in parenthesis.

	Retaliation	Non-retaliation	Total
Prosocial	94 (11%)	482 (56%)	576 (67%)
Antisocial	67 (8%)	219 (25%)	286 (33%)
Total	161 (19%)	701 (81%)	862 (100%)

Table 3.17. Frequency of punishments in Peru, depending on whether they are retaliatory (following punishment in previous round) and prosocial or antisocial. Overall percentage in parenthesis.

	Retaliation	Non-retaliation	Total
Prosocial	26 (6%)	224 (53%)	250 (59%)
Antisocial	31 (7%)	140 (34%)	171 (41%)
Total	57 (13%)	364 (87%)	421 (100%)

Table 3.18. Censored tobit regression of the determinants of prosocial punishment given.

Prosocial punishment given	(1) Total	(2) Brazil	(3) Indonesia	(4) Peru
First order free rider (during baseline)	-0.38***	-0.97***	0.01	-0.43*
	(0.12)	(0.28)	(0.14)	(0.24)
Lagged punishments received (1 round)				
Non-retaliatory (#)	0.10*	0.20	0.05	0.18
	(0.06)	(0.21)	(0.06)	(0.11)
Retaliatory (#)	0.10	0.40	0.03	0.14
	(0.10)	(0.26)	(0.12)	(0.22)
Average conversion	0.31***	0.36***	0.61***	0.01
	(0.08)	(0.12)	(0.13)	(0.11)
Deforestation capacity				
Low capacity	0.28	-0.55	0.57**	0.28
	(0.20)	(0.39)	(0.27)	(0.29)
High capacity	0.39*	0.47	0.10	0.66**
	(0.20)	(0.34)	(0.24)	(0.32)
Round dummy (1 to 6)	-0.07**	-0.06	-0.10**	-0.00
	(0.03)	(0.05)	(0.04)	(0.05)
Constant	-1.37***	-1.51	-0.23	-0.96*
	(0.53)	(1.01)	(0.51)	(0.53)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual level covariates	Yes	Yes	Yes	Yes
Random effects	Yes	Yes	Yes	Yes
Observations	3600	1200	1200	1200
Log likelihood	-2153.56	-590.83	-946.79	-549.95
$\chi 2$	195.46	380.66	187.74	434.31
p-value	0.00	0.00	0.00	0.00

Note: Random effects are at the individual and experimental session level. Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.19. Double censored tobit regression of the determinants of antisocial punishment given.

Antisocial punishment given	(1) Total	(2) Brazil	(3) Indonesia	(4) Peru
First order free rider (during baseline)	0.60***	-0.06	1.16***	0.54*
	(0.18)	(0.31)	(0.25)	(0.32)
Lagged punishments received (1 round)				
Non-retaliatory (#)	0.40***	0.90***	0.19***	0.68***
	(0.08)	(0.15)	(0.07)	(0.18)
Retaliatory (#)	0.42***	0.88***	0.19	0.56**
	(0.13)	(0.20)	(0.17)	(0.23)
Average conversion	0.15*	0.30**	0.50***	-0.02
	(0.09)	(0.12)	(0.15)	(0.14)
Deforestation capacity				
Low capacity	0.64***	0.03	0.65**	0.83***
	(0.21)	(0.32)	(0.29)	(0.29)
High capacity	0.42**	0.43	-0.05	0.63*
	(0.20)	(0.26)	(0.30)	(0.32)
Round dummy (1 to 6)	-0.04	0.06	-0.04	-0.11*
	(0.04)	(0.08)	(0.06)	(0.06)
Constant	-3.44***	-4.27***	-3.84***	-2.24**
	(0.69)	(1.05)	(0.88)	(0.87)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual level covariates	Yes	Yes	Yes	Yes
Random effects	Yes	Yes	Yes	Yes
Observations	3600	1200	1200	1200
Log likelihood	-1419.03	-393.92	-549.22	-431.33
$\chi 2$	1744.14	468.80	1728.64	2019.29
p-value	0.00	0.00	0.00	0.00

Note: Random effects are at the individual and experimental session level. Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.20. Poisson regressions of received punishments, for the total sample and by country.

Dependent variable: total punishment received	(1)	(2)	(3)	(4)
Main independent variable				
Deviation from group average	0.36***	0.24***		
	(0.05)	(0.05)		
Deviation from group average^2	0.04***	0.03***		
	(0.01)	(0.01)		
Deforestation (# of plots)			0.56***	0.50***
			(0.09)	(0.10)
Deforestation (# of plots) ^2			-0.02**	-0.03***
			(0.01)	(0.01)
Lagged punishment given				
Retaliatory (#)	0.11**	0.10*	0.10**	0.09*
	(0.06)	(0.06)	(0.05)	(0.05)
Non-retaliatory (#)	0.11**	0.12***	0.11**	0.12***
	(0.04)	(0.04)	(0.04)	(0.04)
Inequality	0.43**	0.35	0.69***	0.32
	(0.20)	(0.22)	(0.25)	(0.21)
Interaction terms				
Inequality # Deviation	-0.16***			
	(0.06)			
Indonesia # Deviation		0.12*		
		(0.08)		
Peru # Deviation		-0.01		
		(0.07)		
Inequality # Deforestation (# of plots)			-0.13**	
			(0.06)	
Indonesia # Deforestation (# of plots)				0.13**
				(0.06)
Peru # Deforestation (# of plots)				-0.11
				(0.07)
Average conversion	0.13**	0.15**	-0.19**	-0.15**
	(0.06)	(0.06)	(0.08)	(0.07)
Round dummy (1 to 6)	-0.05**	-0.05**	-0.05**	-0.04**
	(0.02)	(0.02)	(0.02)	(0.02)
Indonesia		0.83***		0.65**
		(0.26)		(0.31)
Peru		0.02		0.25
		(0.32)		(0.37)
Constant	-2.44***	-2.30***	-2.69***	-2.32***
	(0.52)	(0.40)	(0.52)	(0.42)
Village fixed effects	Yes	No	Yes	No
Individual covariates	Yes	Yes	Yes	Yes
Observations	3600	3600	3600	3600
Log likelihood	-2351.73	-2373.69	-2355.08	-2368.93
$\chi 2$	595.98	394.56	353.36	386.08
p-value	0.00	0.00	0.00	0.00

Note: Random effects are at the individual and experimental session level. Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

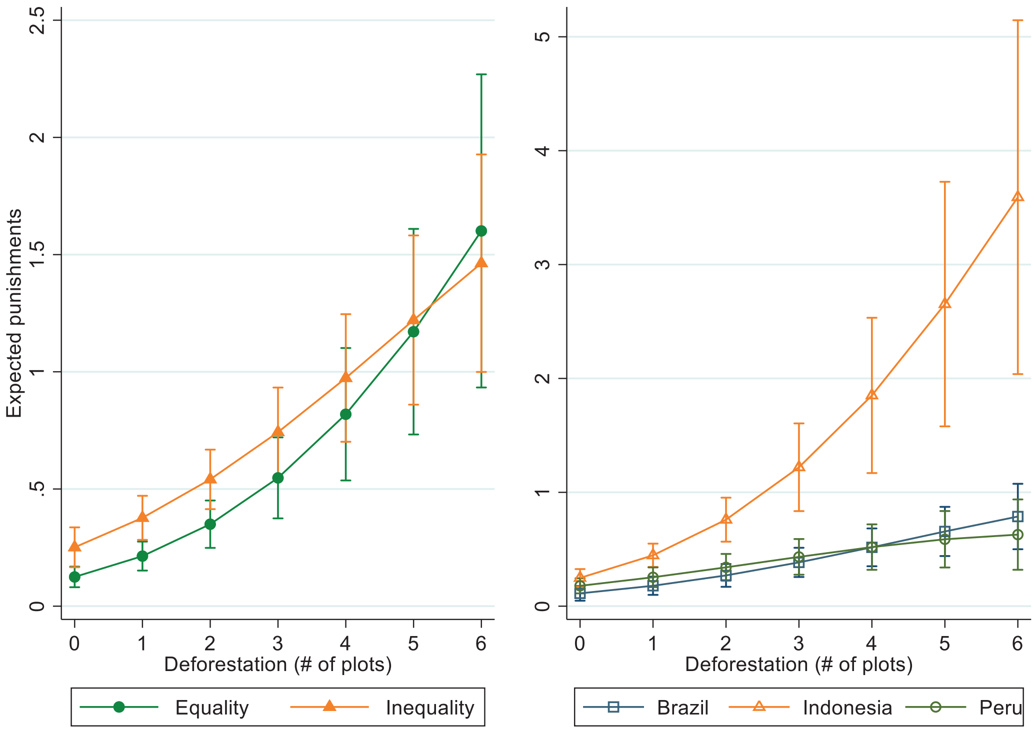


Figure 3.5. Expected punishments received depending on the absolute deforestation level. See table 3.20 for full model specification.

Table 3.21. Impact of lagged punishment received on deforestation levels, considering the deforestation capacity of the player.

	(1)	(2)	(3)	(4)
Change in deforestation	Total	Brazil	Indonesia	Peru
Effect of punishments				
Lagged antisocial punishments	0.264*** (0.080)	0.581*** (0.182)	0.080 (0.087)	0.306** (0.146)
Lagged prosocial punishments	-0.439*** (0.068)	-0.647*** (0.185)	-0.455*** (0.083)	-0.183 (0.161)
Deforestation capacity				
Low	-0.039 (0.035)	-0.110* (0.061)	-0.025 (0.053)	0.032 (0.056)
High	0.015 (0.046)	-0.026 (0.095)	0.115 (0.081)	0.048 (0.053)
Interaction terms				
Low # Lagged antisocial punishments	-0.133 (0.105)	-0.551** (0.255)	-0.026 (0.128)	-0.085 (0.188)
High # Lagged antisocial punishments	-0.013 (0.126)	-0.154 (0.348)	-0.147 (0.112)	0.132 (0.245)
Low # Lagged prosocial punishments	0.106 (0.090)	0.211 (0.231)	0.107 (0.116)	-0.080 (0.195)
High # Lagged prosocial punishments	-0.008 (0.113)	-0.031 (0.277)	-0.034 (0.150)	-0.141 (0.218)
First order free rider	0.088*** (0.028)	0.068 (0.069)	0.144*** (0.046)	0.037 (0.031)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual level covariates	Yes	Yes	Yes	Yes
Observations	3960	1320	1320	1320
Log Likelihood	-6742.798	-2516.440	-1929.547	-2135.283
$\chi 2$	241.380	427.965	1218.872	667.791
p-value	0.000	0.000	0.000	0.000

Note: Random effects are at the individual and experimental session level. Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.22. Impact of lagged punishment on the change in the deviation from the group average.

Dependent variable: Δ Deviation from group average	(1) Total	(2) Brazilian site	(3) Indonesian site	(4) Peruvian site
Lagged punishments received				
Antisocial (#)	0.175***	0.394***	0.031	0.180**
	(0.050)	(0.131)	(0.048)	(0.091)
Prosocial (#)	-0.388***	-0.543***	-0.347***	-0.388***
	(0.069)	(0.195)	(0.083)	(0.117)
First order free rider	0.053*	0.081	0.119**	-0.019
	(0.032)	(0.082)	(0.049)	(0.035)
Interaction terms				
FO free rider # Antisocial	0.263**	0.098	0.280**	0.369*
	(0.108)	(0.334)	(0.139)	(0.190)
FO free rider # Prosocial	0.011	0.000	-0.059	0.165
	(0.072)	(0.247)	(0.083)	(0.107)
Inequality	0.013	0.004	0.043	-0.001
	(0.013)	(0.024)	(0.032)	(0.020)
Round dummy (1 to 6)	-0.006**	-0.006	-0.015***	0.002
	(0.002)	(0.005)	(0.005)	(0.003)
Constant	-0.160**	-0.249**	0.022	-0.130*
	(0.065)	(0.121)	(0.118)	(0.075)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual level covariates	Yes	Yes	Yes	Yes
Observations	8280	2760	2760	2760
Log Likelihood	-6351.547	-2407.425	-1788.251	-1984.975
χ2	230.276	111.477	591.875	566.330
p-value	0.000	0.000	0.000	0.000

Note: Random effects are at the individual and experimental session level. Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4 Paper III

"It is my responsibility as a scientist to ascertain what problems individuals are trying to solve and what factors help or hinder them in these efforts". (Ostrom 1999)

Land use and livelihood impacts of two collective incentive-based conservation interventions in Ucayali, Peru

Julia Naime[1,2], Arild Angelsen[1,2], Erin O. Sills[2,3], Dawn Rodriguez-Ward[2,4]

[1] School of Economics and Business, Norwegian University of Life Sciences (NMBU), PO Box 5003, 1432, Ås, Norway

[2] Center for International Forestry Research (CIFOR), Bogor, Indonesia

[3] Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, United States

[4]Genetic Engineering and Society Center, North Carolina State University, Raleigh, NC, United States

Abstract

Interventions simultaneously addressing poverty alleviation and conservation goals are commonly used to reduce tropical deforestation. We examine the determinants of participation in two interventions that offer collective incentives in Ucayali, Peru. Both interventions are at an early stage of implementation. The first is a local REDD+ project led by an NGO, and the second is the Peruvian's government National Forest Conservation Program (NFCP). We focus on evaluating land use and livelihoods impacts using quasi-experimental methods as well as self-reflexive evaluations (perceptions). We find that households' within village participation is negatively associated with agricultural income and positively associated with market access and previous experiences with conservation or development interventions. We find limited evidence of early-stage impacts (i.e., anticipation effects) on livelihoods or land use. Self-reflexive evaluations, on the other hand, indicate most participating households perceived that both interventions had affected their use of natural resources. A total of 81% of the NFCP participating households indicate an overall positive effect on wellbeing, while only 36% of households perceived the REDD+ project to a have positive impact. The lower perceived positive impacts of the REDD+ project is attributed to design and implementation factors, including delayed payments, low transparency, and limited participation in program development. The study brings forward the importance of considering self-reflexive evaluations to identify possible 'intangible' effects on wellbeing.

JEL classification: Q18; Q23; Q56

Keywords: Deforestation; REDD+; Payments for Ecosystem Services; Rural livelihoods; Poverty

4.1 Introduction

Converting forests to agricultural land is an economic necessity for many rural households in tropical developing countries. To simultaneously address deforestation and rural poverty challenges, two commonly used incentive-based policies are integrated conservation and development projects (ICDPs) and payments for ecosystem services (PES) (Börner et al., 2020). In theory, PES promised to be more effective and cost-efficient than ICDPs to address both environmental and poverty alleviation goals (Ferraro, 2001; Wunder et al., 2020). In practice however, the distinction between PES and ICDPs is often blurry (Wunder et al., 2020), and many PES labelled projects have ICDP components (Sunderlin et al., 2018). For instance, the compensation often comes as in-kind transfers for investment in public goods at the community level or as assets to promote alternative activities, not always covering the opportunity costs of conservation. Furthermore, the conditionality component of PES is seldom adequately enforced and monitored (Wunder et al., 2018), and payments are not directly dependent on the households' individual contribution to the provision of ecosystem services, making many PES projects similar, in practice, to ICDP projects[14].

A growing literature has examined and synthesized the impacts of incentive-based initiatives in terms of both environmental (Pattanayak et al., 2010; Alix-Garcia and Wolff, 2014; Puri et al., 2016; Börner et al., 2020; Wunder et al., 2020) and well-being outcomes (Samii et al., 2014; Duchelle et al., 2018; Liu and Kontoleon, 2018; Pirard et al., 2019) finding in general positive, although small, effects on both outcomes. The participation in and the impact of incentive-based interventions depends on the design, the characteristics of the participating households, and the context in which policies are implemented (Brouwer et al., 2011; Börner et al., 2017; Börner et al., 2020). One essential design issue of incentive-based interventions is whether to deliver the incentives collectively (to a group) or individually (Engel, 2016). Even though collective payments are subject to free-riding and thus potentially lower effectiveness (Gatiso et al., 2018; Hayes et al., 2019; Ngoma et al., 2020), they are more appropriate – and even necessary – when land is managed or owned collectively. This article examines two collective PES conservation initiatives in Ucayali, Peru. We consider a "user financed PES" and a "government financed PES" (Engel et al., 2008; Pirard et al., 2019) at early stages of implementation: the first is a subnational REDD+ project run by a Peruvian NGO, and the second is a government PES-ICDP program, the National Forest Conservation Program (NFCP).

We contribute to the impact evaluation literature of conservation initiatives in three ways. First, due to the early stage of implementation of the interventions, we evaluate how *expectations* of future income and investment affects environmental and livelihoods outcomes, a topic which has been left unexplored in the context of PES. While there is an increasing examination of post-program PES impacts (e.g., Pagiola et al., 2016; Hayes et

[14] In the case of REDD+ projects, this means that the compensation is not tied to higher carbon stocks and/or sinks, see West et al. (2020) an example.

al., 2021), there has been less emphasis on behavioural changes before treatment is implemented. This is in spite the fact that theoretical and empirical analyses suggest that expectations at the start or during conservation projects can mobilize actors and resources (Harstad, 2016; Massarella et al., 2018). For instance, expected future income affects investment decisions of households (Aggarwal and Brockington, 2020), and postponed or delayed payments can promote the consumption of conservation goods (Harstad, 2016). Such anticipation effects occur when there are changes in behaviour before treatment implementation (Malani and Reif, 2015), and they are particularly important in the context of PES, as they can modify the reference levels and thus change project impacts and additionality.

Second, we contribute to the relatively scarce literature evaluating household-level livelihood and land use outcomes of collective agreements (Hayes et al., 2019). Both the NFCP and REDD+ agreements are collective in nature as performance is measured at the village level, but the degree of participation varies across households within villages. We first examine the determinants of program participation, and then evaluate whether the potential anticipation effects differ between households who are actively participating and benefiting from the program and those who are not. Understanding participation and distinguishing between participating and non-participating households impacts is important given that a central concern of collective agreements is the equitable distribution of costs and benefits at the community level (Hayes et al., 2019).

Third, we estimate and compare impacts using quasi-experimental evaluation based on three years of survey data collection (2012, 2014, and 2018), as well as self-reflexive responses (i.e., perceptions) based on open-ended survey questions (Schreckenberg et al., 2010). The two interventions differ in design and implementation characteristics that can affect households' wellbeing and land use outcomes in multiple ways. Both methods provide complementary information and thus allow a more comprehensive understanding of program impacts. The rest of the paper is structured as follows. In section 4.2, we first describe the interventions and their stage of implementation, before proceeding to the theory of change about the expected program impacts. In section 4.3, we present the study site, quasi-experimental methods and definitions of self-reflexive evaluation. In section 4 we present the results about program participation and impact. In section 4.5 we discuss the results, before concluding in section 4.6.

4.2 The interventions and the theory of change.

4.2.1 Description of the interventions.

The first intervention evaluated is the user-financed REDD+ project undertaken by the Peruvian NGO AIDER. The project was launched in 2012 in seven indigenous communities. Project interventions have promoted activities such as sustainable extraction of non-timber and timber products as one way to reduce forest carbon emissions and generate carbon credits (Rodriguez-Ward and Paredes, 2014). The project did not significantly affect the declining forest income during the years 2012-2014 (Solis et al., 2021). In 2015, the seven communities, together with AIDER, obtained certification for reduced deforestation from the VCS and CCB standards[15]. The sale of carbon credits took some time, with negotiations between Althelia Climate Fund (now part of Mirova Natural Capital), AIDER, and the communities starting in 2017 and culminating with the purchase of 222,400 Verified Carbon Units (Althelia, 2018). It was agreed that part of the benefits from the sale of carbon credits would be invested into productive alternatives in the participating villages.

At the time of our survey, in September-October 2018, AIDER had organized REDD+ workshops in the participating villages to explain the agreement with Althelia and to elaborate a pre-investment plan. The pre-investment plan described the investment activities to be implemented in each village by the participating households. The communities were expecting to receive household level distribution of assets, and the purchase of communal equipment to prepare timber and agroforestry plantations (e.g., trucks) had started. The nurseries for timber and agroforestry plantations were initiated in 2019, with the full establishment of the plots in 2020. Examining anticipation effects is particularly relevant given the 3-year delay between the carbon certification and the benefit distribution, as expectations of the REDD+ project could have affected land use and livelihood (i.e., income) outcomes.

The second intervention evaluated is the government financed intervention, the National Forest Conservation Program (NFCP), which started implementation in the area in December 2017, after the Peruvian environmental ministry (MINAM) signed three-year agreements with four villages willing to participate in the program[16]. During the three-year collective agreement, each community receives yearly cash funds, equivalent to 10 soles (USD 2.45) per ha enrolled, conditional on the undertaking of activities specified in a predefined investment plan. The investment plan is elaborated by the communities together with government representatives and approved in village assemblies. It has four budget components: social, environmental, productive, and management. The *social* component mainly aims to improve village level infrastructure, such as communication infrastructure or the establishment of a health emergency fund. The *environmental* component provides equipment for the forest monitoring patrols (e.g., GPS, boots, lamps, sleeping bags, first aid kits, compass, and in some cases drones). As part of the agreement, participating communities must monitor the village boundaries at least once every three months to detect illegal settlements or illegal logging activities. The *productive* component provides inputs for fishing activities (e.g., fishing nets), carpentry activities (e.g., chainsaws, transportation vehicles), artisanal textile production, ecotourism activities, as well as consumption goods such as small livestock (ducks,

[15]CCB standards are certified by Verra: https://registry.verra.org/app/projectDetail/CCB/1360
[16] Nationally, the NFCP began in 2010, and has received donor funds for its implementation since 2014, see Joint Declaration of Intent between Germany, Norway and Peru: https://www.regjeringen.no/contentassets/b324ccc0cf88419fab88f2f4c7101f20/declarationofintentperu.pdf

chickens and pigs). Finally, the *management* component is reserved for the transportation and administrative costs of implementing the three-year investment plan.

The NFCP also includes educational workshops about legal requirements and steps to regularize the extraction of forest products. In 2018, participating communities were working on obtaining government approved management plans to extract natural resources from the forest. Nationally, the programme has had small positive effects on forest cover (Giudice et al., 2019; Giudice and Börner, 2021), but there is no evaluation of its livelihood impact at the household level yet. By the end of 2018, communities had received capacity-building workshops and equipment to increase forest monitoring, as well as in-kind benefits such as textile assets (thread and cloth) for handicraft production, construction materials to build a tourist lodge, as well as small livestock assets and fishing equipment. In 2019, communities obtained approved management plans for timber extraction activities, with reported sales in 2020. Section 2 in the Appendix C provides a timeline summarizing the two projects.

4.2.2 Research questions and theory of change

Given that the agreements are collective but participation amongst households differs within communities, our first research question is to identify the determinants of participation in the interventions. When households are constrained and cannot exchange all their available capital and labour in the market, participation is guided by internal, shadow prices on labour use. Shadow prices are not directly observable but can be proxied by other household characteristics. For example, variables such as family labour and access to credit influence participation (Pagiola et al., 2005; Jayachandran, 2013; Jack and Jayachandran, 2019).

We expect participation in these interventions to depend on households' opportunity cost of participation, determined by the expected benefits of the program (Weber et al., 2011). In the early phase of implementation that we study, participation in the REDD+ project means attending workshops to develop the pre-investment plan or educational and capacity-building workshops. Participation in the NFCP means attending the educational and capacity-building workshops offered by the program, and/or having received household level assets from the program. We hypothesize that households who have lower levels of labour market integration, lower shadow wages and thus opportunity costs of participation are more likely to participate in either of the projects.

Our second research question concerns the impacts of the program at this early stage of implementation. We develop a causal chain to present our theory of change (ToC) and specify how the interventions can affect forest conservation and livelihoods (Ferraro and Hanauer, 2015; Sills and Jones, 2018). Elaborating a causal chain involves identifying the pathways, or mechanisms, along which an intervention can deliver an impact, and the critical assumptions and conditions that moderate those pathways (Qiu et al., 2018).

To describe the causal chain, in each intervention we identify two main components (Fig. 4.1). The first is an 'information' component, comprised of the workshops to raise awareness, build capacity, and develop investment plans. The second is a 'productive asset' component, involving the provision of community and household level assets. The interventions could have changed the expected income of alternative activities, which would then shape participating households' decisions about forest use and deforestation.

Under imperfect markets, an inflow of resources from the conservation interventions can relax human capital constraints (by providing skills and education), or physical and financial capital constraints such as new equipment and technologies (Groom and Palmer, 2010; Liu and Lan, 2015). These in turn change the relative profitability of alternative sources of income, making the overall outcome on forest use and deforestation uncertain (Angelsen, 1999). For example, relaxing households credit or input constraints could have negative impacts on conservation and increase deforestation (Vosti et al., 2001; Alix-Garcia et al., 2013), while relaxing non-agricultural labour constraints can improve environmental outcomes (Groom et al., 2009).

The direction of the impact on participating households is ambiguous. On the one hand, the awareness workshops, capacity-building trainings, and promises of productive assets, could create expectations about future reduced input and increased output prices for sustainable activities, which increases the expected forest rent as compared to agricultural. Coupled with the increased environmental awareness of the households and increased NGO and government presence, the interventions could have reduced deforestation of the participating households. On the other hand, the provision of assets, together with the fear of increased forest monitoring in the future as the projects continue their implementation, could have increased deforestation and forest extraction activities of the households. Both effects are to be interpreted as 'anticipation effects' in which the behaviour of households changes before full implementation of the treatment.

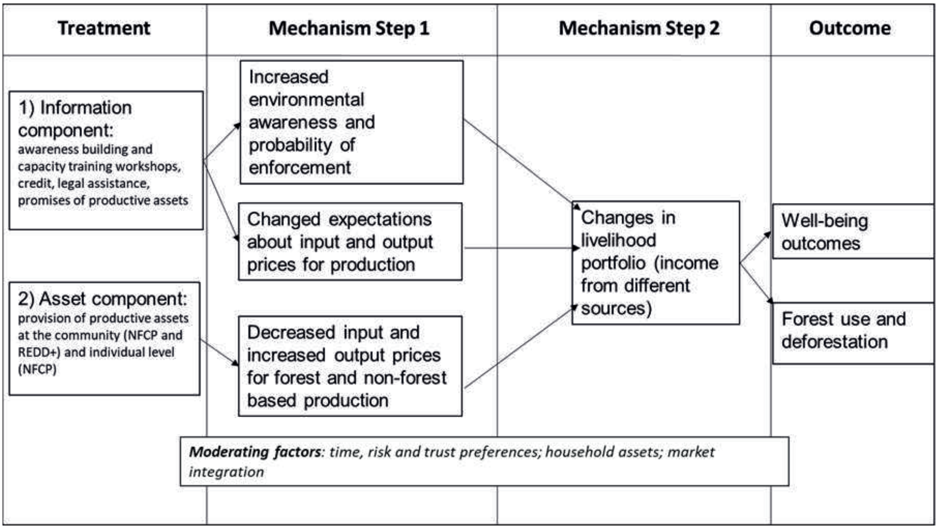


Figure 4.1. Theory of change identifying critical conditions for effectiveness of ICDP-PES interventions.

Furthermore, the collective agreements could also have affected non-participating households in the villages where the projects are being implemented. For example, the 'information' component could spill over to non-participating households and have an impact on their land use or income. Further, increased monitoring and government and/or NGO presence could restrict the forest extraction activities of non-participating households, also affecting their livelihoods. To better understand the intra-community impacts of the collective agreements, we thus distinguish between the impacts on participating households from the impacts on non-participating households. We first evaluate the intention-to-treat (ITT) effect to identify overall program impacts, which is the effect of being in a village where the collective agreement is offered (i.e., it pools the participating and non-participating households). Subsequently, we evaluate separately the Average Treatment Effects on the Treated (ATT), that is, the impact on the participating households, and the average treatment effect on the non-participating households, which we label ATNP.

4.3 Study site and methods

4.3.1 Description of the study site.

The study site comprises eight villages belonging to the Shipibo-Konibo indigenous group, all located in the Ucayali district. Ucayali is one of the two districts with the highest terrestrial carbon (Asner et al., 2014; Csillik et al., 2019) and the highest deforestation rates in Peru (MINAM, 2020). The drivers of deforestation in the area are mostly related to agricultural expansion, with oil palm playing an increasing role (Bennett et al., 2018).

The ecological conditions in which these villages are located, as well as their distance to urban markets, constrains households' primary livelihood strategies to agricultural, fishing, and forest extraction activities (Coomes et al., 2010; Rodriguez-Ward and Paredes, 2014; Porro et al., 2015; Begazo Curie et al., 2021). Participation in the labour market is seasonal and sporadic. Three villages are in seasonally flooded forests with limited potential for agricultural expansion, while five villages are in non-floodable forests and thus more suited for agricultural activities. Farming, fishing, and agricultural activities occur mostly during the dry season, which runs from April to September.

Out of the eight studied villages, two villages received both NFCP and REDD+ projects, two villages participated only in the NFCP program, two villages participated only in the REDD+ project, and in two villages neither project was implemented. We collected village and household level information in the years 2012, 2014, and 2018, surveying approximately 30 randomly selected households in each village. Income data was collected closely following the Poverty and Environment Network (PEN) methodology (Angelsen et al., 2014).

Income questions were separated into seven categories: forest income, fishing income, agricultural income (cash and subsistence), livestock and animal products, wage income, and other income (business, pensions, remittances). For all our measurements, the costs of purchased inputs were deducted from gross values. Values from 2012 and 2014 were adjusted for inflation to 2018 prices. We converted the income from Peruvian *nuevos soles* (PEN) to USD using the exchange rate 1 USD = 4 PEN. We also asked households about their livestock holdings, ownership of durable goods (e.g., phone, tv, solar panel), and their productive assets (e.g., agricultural equipment). Given that land is communally owned in all villages (there is not *de jure* individual property), we elicited the area of agricultural and forest land that is managed or controlled by the household *de facto*.

Table 4.1 indicates main village characteristics depending on which intervention is implemented. Across the villages, we find no significant differences in forest income, wage, assets, or forest land used. Socioeconomic variables such as education, age, households' size and participation in external interventions were not significantly different either. The most important differences are found between the two 'NFCP only' villages and the villages where neither intervention is offered: the 'NFCP only' villages have lower agricultural income, lower agricultural land, and households left a smaller quantity of land fallow in the two preceding years and had lower deforestation on average. This variation is due to the fact that the two 'NFCP only' villages are in areas with floodable forest and thus limited agricultural potential. Our sample is consistent with the national level evidence showing that the NFCP has (thus far) been implemented in communities with relatively lower deforestation rates[17] (Giudice et al., 2019).

A final relevant difference is that the villages where neither intervention was offered (Table 4.1, column 2) have on average a higher percent of cash income, and thus higher

[17] We note that since 2018, the year the data was collected for this study, the NFCP has continued its expansion to communities that are further away from the main urban market of the city of Pucallpa, in a progressive roll out of the program.

market integration. Overall, these characteristics are suggestive of a village level selection bias originating from both the project implementers and local residents: villages with higher market integration and higher deforestation rates might be less interested in welcoming external conservation interventions, and project proponents might be less interested in intervening in villages where opportunity costs of conservation are higher.

Table 4.1. Summary statistics of villages in the baseline year, 2014 (N=244). Standard errors in parenthesis. Same superscripts indicate no significant difference in means between groups (Bonferroni tests).

Summary statistics at the village level	(1) Both interventions		(2) No intervention		(3) NFCP only		(4) REDD only	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Total deforestation (has/2yrs)	.61[a,b]	.17	.74[a]	.12	.2[b]	.06	.44[a,b]	.07
Land left fallow (has/2yrs)	.67[a,b]	.18	.96[a]	.25	.27[b]	.07	.39[a,b]	.09
Agricultural income (USD/yr)	1077.79[a,b]	161.81	1252.26[a]	109.72	262[b]	36.78	798.74[b]	69.04
Forest income (USD/yr)	518.28	69.77	326.39	48.58	438.97	94.25	504.28	102.55
Fishing income (USD/yr)	1478.64[a]	195.02	871.09[b]	128.12	1347.94[a,b]	194.03	1227.12[a,b]	160.52
Wage income (USD/yr)	566.58	147.19	663.64	160.56	659.63	185.94	632.38	136.88
Other income(USD/yr)	226.29[a,b]	70.2	451.73[a]	89.84	146.95[b]	50.96	391.69[a,b]	92.73
Cash income (total income share)	.45[a]	.03	.56[b]	.03	.34[a]	.04	.35[a]	.03
Assets (USD)	333	49.76	376.74	45.38	333.93	105.36	518.12	55.86
Livestock assets (USD)	22.08	6.36	70.82	33.06	19.26	5.21	46.84	8.56
Agricultural land (has)	2.16[b]	.31	2.76[b]	.36	.86[a]	.11	1.41[a,b]	.16
Forest land (has)	2.77	1.09	2.69	.37	1.21	.23	2.70	.35
Mean household age	24.78	1.32	23.84	1.42	23.63	1.51	23.23	1.52
Mean household education (yr)	6.84	.29	6.58	.29	5.93	.34	6.57	.29
Household size (#)	5.55	.27	5.88	.26	5.52	.3	6.19	.29
Household participation in programs (share of offered programs)	.44	.04	.32	.04	.39	.05	.33	.04

4.3.2 Difference in difference with matching

To calculate the average ITT, ATT, and ATNP of each intervention we conducted differences-in-differences (DiD) regressions, with inverse propensity score weighting in a sample trimmed of any observations outside the range of common support defined by the propensity score (Hirano et al., 2003; Abadie and Imbens, 2016). PSM allows to control for any observable selection bias of the project participants and mimics randomization by ensuring that there are no observable differences between the treated and the control units (Ferraro and Hanauer, 2014). Unlike normal matching, PSM addresses the dimensionality problem of having too many variables to match to (Rosenbaum and Rubin, 1983; Abadie and Imbens, 2016).

The DiD framework allows, in turn, to control for any selection bias due to unobservable variables that do not vary over time. Such variables could be, assuming stable preferences, the risk and time preferences of households, or their entrepreneurial skills, or village characteristics that affect selection into the program. A key assumption of DiD is the parallel trends assumption, which holds that the selection bias between treated and control units is time invariant, i.e., the fixed, unobservable differences can be controlled for by focusing on difference in outcomes over time (Khandker et al., 2009).

Our sample for the quasi-experimental evaluation is comprised of a total of 187 households who were surveyed in both 2014 and 2018 (Table 2). We take as our baseline the year 2014, and test for parallel trends in 2012-2014. From the 2014 sample, 27 households were not surveyed in 2012, and thus could not be included in the parallel trends test[18]. In villages where only REDD+ was offered, 24 out of 49 households participated in the intervention. In villages where only the NFCP was offered, 26 out of 33 households participated. In villages were both REDD+ and NFCP projects were offered, a total of 17 households were participating in both the REDD+ and NFCP projects, 12 were participating only in REDD+, 8 were participating only in NFCP, and the remaining 10 households did not participate in either program.

Table 4.2. Description of the sample for the quasi-experimental impact evaluation, by type of village and year. The number of participating households are indicated in parenthesis. The sample comprises all the households that were interviewed in, at least, the years 2014 and 2018.

Year	Non-intervention villages	REDD+ villages	NFCP villages	REDD+ and NFCP villages	Total
2012	46	46	26	42	160
2014	58	49	33	47	187
2018	58	49 (24)	33 (26)	47 (17 both, 12 in NFCP, 8 in REDD, 10 in none)	187

We consider the impact of each program separately, as the selection of the intervention villages was conducted independently by the NGO AIDER and the NFCP implementers. A critical question of the matching procedure was how to select our control group. We chose the procedure that could maximize the number of control observations for each treatment, as matching using only the 58 ‘non-intervention villages’ observations reduced the matching quality and resulted in lower common support between treatment and control observations. This implies that the control sample of the REDD+ program are the 58 households of the non-intervention villages plus the 33 households of the NFCP villages (91 in total). In turn, the control sample of the NFCP intervention are the 58 households of non-intervention villages plus the 49 households of the REDD+ only

[18] For an analysis of attrition, see Appendix C Table 4.7.

villages (107 in total). In all of our analysis we control for the fact that households might be participating in both of the programs.

To calculate the ITT, we use both participants and non-participants of the programs, which involves a total of 96 treated observations in the case of the REDD+ program and 80 observations in the case of the NFCP program. Our control sample for the REDD+ intervention involves the 91 households in the non-REDD+ villages, and the control sample for the NFCP interventions involves the 107 households in villages where that intervention is not offered. All control observations are weighted by inverse propensity scores.

To calculate the ATT, we considered households participating in the programs and compared them to the control sample. For the REDD+ program, we estimate the ATT for the 49 participating households (24 +17 + 8) by comparing them with the 91 control households. For the NFCP, we estimate the ATT for the 55 participating households (26+17+12) by comparing them with the 107 households in the villages where the NFCP is not offered.

Finally, to calculate the ATNP, we match and compare the non-participating households in intervention villages to households in control villages. This yields 47 (25+22) non-participating observations for the REDD+ treatment, and 25 (18+7) for the NFCP program. Thus, for all the effect estimations of each intervention, the control sample is the same, but we change the pool of "treatment" observations.

In the matching process we included variables related to both treatment assignment as well as the outcome (Stuart, 2010), as done in previous impact evaluation studies of conservation interventions (Simonet et al., 2018; Montoya-Zumaeta et al., 2019; Solis et al., 2021). We use the pre-treatment, baseline year 2014 to conduct the matching procedure. Even though the REDD+ project started working in 2012, 2014 is a good baseline year to evaluate anticipation effects, recalling that the VCS-CCB certification was obtained in 2015. The 2014 matching variables were stated deforestation, land left fallow, farm income, environmental income, fishing income, wages and other income, share of cash income, total value of assets, agricultural and forest landholdings, household size, mean household age, mean household education, and the proportion of previous household participations in forest interventions from the total offered in the village.

The estimations of the treatment effects can vary depending on the matching procedure (Khandker et al., 2009), therefore we evaluated multiple matching algorithms (see Appendix C, section 3-5). We used the absolute standardized difference of the means in the treated and (matched) non-treated group (Rubin's B) and the ratio of treated to (matched) non-treated variances (Rubin's R) as indicators of quality matching. Rubin (2001) recommends that B be less than 25 and that R be between 0.5 and 2 for the samples to be considered sufficiently balanced. These indicators had been used for the impact evaluation studies of conservation interventions (Chervier and Costedoat, 2017).

For the REDD+ project, matching satisfactorily reduced Rubin's B and R to 20.9 and 1.05 in the case of the ATT, to 27 (Rubins B) and 0.72 (Rubins R) in the case of the ATNP, and to 21.6 and 1.06 in the case of the ITT. For the NFCP, the matching procedure only reduced Rubin's B and R to 60.2 and 0.92 in the case of the ATT, to 24.5 and 2.34 in the case of the ATNP, and to 27.7 and 1.65 in the case of the ITT. The unsatisfactory matching of the NFCP' is attributed to the stark differences in livelihood strategies between villages (cf. Table 4.1). We therefore conducted the estimations with regression adjustment by including in the regression all the matching variables which had a remaining standardized bias higher than 10% (Rubin, 1973; Nguyen et al., 2017).

We found no significantly different pre-intervention trends between REDD+ participating and control observations for any outcome variable, which lends support to the parallel trends assumption. Nevertheless, the REDD+ non-participating households have higher forest land that the control observations. For the NFCP program, there is no evidence of parallel trends in forest land and fallow land: households in NFCP village have less forest and fallow land than non-NFCP villages. In the results section we only show coefficients for estimations based on comparisons that followed parallel trends in 2012-2014. For the analysis of the parallel trends' assumptions, see Appendix C, section 6.

4.3.3 Definition of variables and econometric models.

To evaluate program participation, we use logit regressions, defined as follows:

$$Pr(P = 1) = F(\alpha_0 + \alpha_i X_i + \epsilon_i) \qquad (4.1)$$

Where Pr(P=1) is the probability of participating in the NPCF and REDD+ programs, $F(Z) = \frac{e^Z}{1+e^Z}$ is the cumulative logit function, and X_i is the vector of socioeconomic covariates for each household. To avoid simultaneity between dependent and independent variables, we use the observations of the baseline year (2014) to examine program participation. We examine the impact of market integration on participation using two different indicators: (i) the income from wages and business, and (ii) the percent of income received in cash (i.e., sold) from the total environmental and agricultural incomes. We could expect opposite effects from each indicator. On the one hand, higher access to the labour market reduces participation because of higher shadow wage. In turn, if market integration is proxied with percent of cash income, as an indicator of market access, we could expect participation to increase if it increases expected returns from the program's activities.

We examined impact on four land use variables: (i) deforestation, (ii) land left fallow, (iii) agricultural land used, and (iv) forest land used, all reported by household respondents. Although all the variables are self-reported measures and thus are subject to recall and measurement problems, they are unlikely to generate bias if they are random in nature. Furthermore, our confidence on the validity of the land use variables is backed by the fact that the self-reported measures of land use and deforestation have been found to coincide with satellite images (Simonet et al., 2018).

To examine impact on income, we separated income by economic activity since this is a useful avenue to identify alternative mechanisms of impact of an intervention (Le Velly and Dutilly, 2016). We adopt a classification based on the use of natural resources, classifying income into (i) farm income, referring to the income coming from agricultural and livestock activities, (ii) environmental income refers to income from forest and fishing, and (iii) 'other' income refers to the income obtained from wages, business, and miscellaneous.

For all the regressions, we dropped observations that did not satisfy the common support assumptions and eliminated observations of the dependent variables which were below or above the 5% and 95% percentile to avoid outlier sensitivity. All our regressions are conducted using inverse propensity score weighting with regression adjustment. Our main econometric model was:

$$\Delta Y_i = \beta_0 + \beta_1 D1 + \beta_2 D2 + \beta_i X_i + \epsilon_i \qquad (4.2)$$

Where ΔY_i indicates the difference in outcome variables from the years 2014-2018, $D1$ indicates the treatment being evaluated (either REDD+ or NFCP), and $D2$ is the second intervention (either REDD+ or NFCP) introduced as a control variable, given that some households participate in both programs, and X_i is a vector of covariates which have a remaining standardized bias higher than 0.1. The average ATT, ATNP, and ITT of the program is indicated by the coefficient β_1. The coefficient β_2 cannot be interpreted casually, because the sample is not weighted to achieve balance on D2. The term ϵ_i represents the idiosyncratic error term.

4.3.4 Self-reflexive evaluation of the interventions

Experimental and quasi-experimental methods are good for inferring causal attribution and extracting the average treatment effects in a sample or population. However, the program's *perceived* benefits may be just as important in shaping household responses to the program including whether they continue to participate and they can provide additional insights regarding procedural aspects of program implementation. We thus asked the participating households of each program whether and how the program had affected their land use activities, with the question: *"Has the intervention affected the way you use land and other natural resources (e.g., agricultural or livestock practices, use of forests and forest resources, forest clearing)?"*.

We also asked whether the intervention had positively or negatively affected their overall wellbeing, with the question: "*What is your evaluation of the effect of the intervention on the wellbeing of your household?"*. We kept a broad definition of land use or wellbeing to allow households to state what was most important for them when evaluating the intervention. We used logit regressions to examine if the perceptions about positive or negative impact were correlated with socioeconomic characteristics. A limitation of our survey is that only participating households were asked for their self-reflexive evaluations. To mitigate this potential source of bias, we corroborate our findings with results from focus group discussions held in each village.

4.4 Results

4.4.1 Participation in the intervention

The logit regressions indicate two significant variables that determine participation of both interventions (Table 4.3). First, having participated in an external intervention in the past increases the odds of participation by 50% on average in the case of the REDD+ project, and practically doubles it for the NFCP project. Previous participation might reflect households' "entrepreneurial spirit" or the trust in external interventions. Second, households with higher agricultural income are less likely to participate in the REDD+ and NFCP projects. In particular, a 1% increase in income decreases the odds of participation by 0.7 and 0.3 for the NFCP and REDD+ program respectively. This is consistent with the hypothesis that households with higher opportunity costs are less likely to participate in the program.

The indicator of labour market integration is not significant, while the indicator for market access has a positive effect on NFCP participation. Further, engaging in deforestation activities as well as having a greater amount of forest land is negatively related to participation in the NFCP program, but not in the REDD+ project. Thus, overall, participation in the NFCP is most likely for households with less agricultural income and higher market access. The fact that there are fewer significant variables in the case of the REDD+ project (and an overall lower model significance, with $p>0.1$) is likely explained by the early stage of implementation of the project, and that the treatment and participation definition are comparatively weak, due to lack of disbursement of funds or assets at the individual level.

Table 4.3. Logit model of determinants of household level program participation in the REDD+ initiative and the NFCP.

Dependent variable: program take up rate	(1) Participation in NFCP	(2) Participation in REDD+
Land use variables		
Deforestation dummy	-1.63** (0.71)	0.23 (0.43)
Agricultural land (log)	0.97 (0.77)	-0.09 (0.40)
Forest land (log)	-0.75* (0.40)	0.08 (0.30)
Income variables		
Agricultural income (log)	-0.69*** (0.22)	-0.27* (0.15)
Forest income (log)	-0.03 (0.11)	0.14 (0.12)
Fish income (log)	0.18 (0.15)	0.10 (0.16)
Livestock income (log)	0.14 (0.11)	0.07 (0.07)
Assets		
Total value assets (log)	-0.08 (0.13)	0.11 (0.11)
Demographic variables		
Mean education of hh members	-0.08 (0.13)	0.11 (0.11)
Mean age of hh members	-0.05 (0.04)	0.02 (0.03)
Household size	0.18 (0.18)	-0.01 (0.12)
Years hh lived in village	0.03* (0.02)	-0.01 (0.02)
Number of programs previously participated	0.67*** (0.22)	0.39*** (0.13)
Indicators of market integration		
Wage and other cash income (log)	-0.08 (0.09)	0.05 (0.08)
Market access: Percent cash income	1.72** (0.83)	-0.49 (0.83)
Constant	1.75 (2.96)	-1.45 (1.97)
Observations	114	123
Adjusted R^2	0.37	0.13
Village dummies	Yes	Yes
$\chi 2$	40.51	24.71
p-value	0.00	0.13

Note: Robust standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Variables with (log) are transformed using the Inverse Hyperbolic Sine (IHS) transformation.

4.4.2 Quasi experimental evaluation of anticipation effects of the interventions.

The quasi-experimental analysis indicates that the REDD+ program has not had any significant impact on participating households (Table 4.4). The analysis of the impacts of the NFCP indicates, in contrast, significant impacts on participating households, in the amount of forest land used by 0.8 ha. We note however, that the parallel trends analysis indicates that NFCP households left a larger quantity of land fallow in the years 2012-2014, which could explain the increase in forest land used and managed during 2014-2018. Furthermore, although we find significant effects on the amount of forest land used by REDD+ non-participating households (see Appendix C, Section 7, Table 4.22), the parallel trends assumptions does not hold (Appendix C, Section 6).

Taken together, the results suggest that the NGO presence could be having a cumulative effect (from the years 2012) of increasing the forest land used, but not as a result of the expectations of future program implementation (i.e., anticipation effects). We find no significant impacts on any of the income variables of the participating households, and there is no significant effect on non-participating households either. Overall, we show that the programs are thus not differentially affecting the incomes of participating and non-participating households, and that there are no negative 'anticipation' effects that increase deforestation or agricultural land.

Table 4.4. Average treatment effects on the participants (ATT) , effect on the non-participants (ATNP), and intention to treat (ITT) effects of the REDD+ and NFCP projects, on land use and income variables. Only the coefficients that satisfy the parallel trends assumption are included. See Appendix C, section 7 for full model specification and results.

	(1) Ag. Land (ha)	(2) Forest land managed (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
ATT (participants)							
REDD+	0.16	0.13	-0.07	0.14	86.51	229.62	-239.58
	(0.12)	(0.29)	(0.23)	(0.16)	(86.17)	(207.53)	(149.83)
NFCP	-0.43	0.82^{*}	-	-0.03	53.88	237.70	102.88
	(0.36)	(0.36)		(0.16)	(117.67)	(220.20)	(151.04)
ATNP (non-participants)							
REDD+	0.35	-	0.27	0.22	159.26	78.70	72.52
	(0.27)		(0.21)	(0.13)	(99.19)	(184.17)	(188.35)
NFCP	0.33	-	-	-0.09	-24.98	-48.95	-36.33
	(0.25)			(0.15)	(121.24)	(350.05)	(238.44)
Intention to treat (ITT)							
REDD+	0.11	0.17	0.08	0.20	57.54	8.78	18.20
	(0.14)	(0.19)	(0.10)	(0.12)	(104.48)	(102.68)	(107.41)
NFCP	0.02	-	-0.20	-0.12	25.70	112.74	222.25
	(0.30)		(0.11)	(0.13)	(105.31)	(124.87)	(206.28)

Note: Clustered standard errors at the village level in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.4.3 Self-reflexive evaluations on natural resource use and wellbeing impacts.

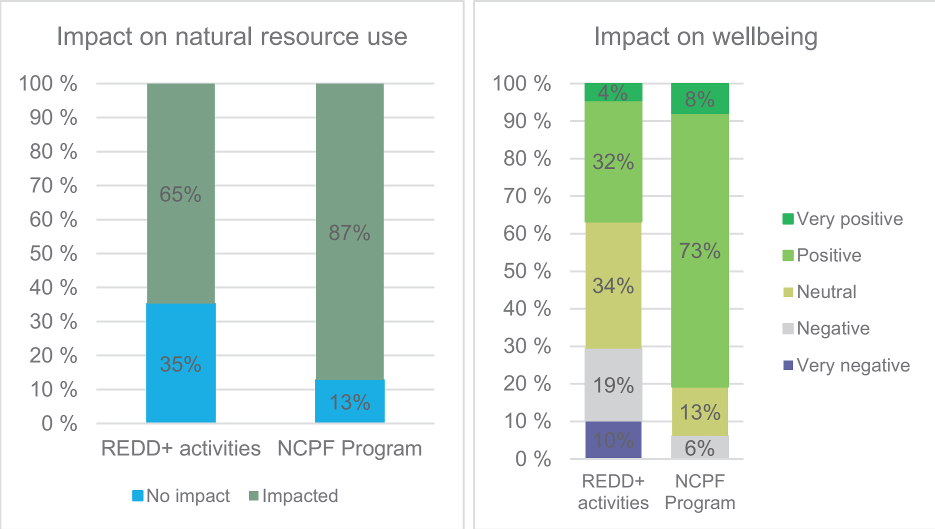
Interestingly, and even though there is limited evidence of program impact from the quasi-experimental analysis, most program participants perceived that the programs had an impact on how they use their resources and their overall wellbeing. Approximately 65% of the households participating in the REDD+ program state that the program is having an impact on their resource use, and 87% NFCP participants indicate an impact (Fig 4.2). Regarding wellbeing, the positive perceptions of participants were more frequent for the NFCP (81%) than for the REDD+ program (36%, Fig. 4.2).

Table 4.6 summarizes the reasons for impact on land use and well-being. The reasons the REDD+ project created dissatisfaction amongst households are the unmet expectations and uncertainty about program content and delivery, including the use of language that is perceived as being too technical for the households. In addition, participants mentioned the lack of participation and transparency in the process designing the pre-investment plan in the REDD+ project. Dissatisfaction from NFCP participants stems from perceptions that the in-kind compensation and the support to conduct monitoring activities was too low. Nevertheless, the participatory nature of the NFCP was well received, and particularly highlighted in village meetings (Table S20). Hence, the difference between the perceived wellbeing impacts of the two projects is partly attributed to procedural aspects during the design and planning of the program.

Remarkably, a large number of participants of both projects mentioned positive effects in terms of improved forest monitoring and management (Table 4.6). Households also mentioned the goods and equipment provided at the village level of the REDD+ project. Such items included receiving a transportation vehicle, chainsaws for timber extraction and processing, and material for improved forest monitoring. Participants also recognized that the REDD+ capacity training workshops allowed them to improve their forest management practices, in, for example, improved secondary forest management for the extraction of "bolaina" (*Guazuma crinita*). NFCP participating households mentioned receiving monitoring equipment, as well as in-kind support for the farm activities (duck, pigs) and non-farm activities (artisanal products, fishing activities).

During the village focus groups, only one village classified the REDD+ project as being overall positive, recognizing its impacts in terms of forest monitoring and improved timber extraction, while two villages assigned a negative effect, indicating that they would prefer to receive the money and carbon compensation directly. The four villages which participated in the NFCP indicated that the program has had overall positive effects, as the result of receiving in-kind compensation and current expectations about future revenue streams from fishing, ecotourism and sustainable logging activities (Appendix C, Table 4.26)

Figure 4.2. Perceived impact of natural resource use and wellbeing, for the REDD+ project and the NFCP, on the participating households.



Finally, we examined whether the perceptions about natural resource use and wellbeing impact were correlated to socioeconomic variables (Table 4.5). In general, there are low correlations between the socioeconomic and demographic variables and perceptions: three models out of four are jointly insignificant. Interestingly, the amount of agricultural land is negatively correlated to the probability of perceiving a positive wellbeing impact of the REDD+ project, while higher forest land is significantly related to positive NFCP impact on wellbeing. These results suggest that gains from the REDD+ project are lower for those with more agricultural land, and the gains from the NFCP higher for those with more forest land. Higher income is related to REDD+ positive effects but is not related to any of the other outcome variables.

Table 4.5. Logit models of participating households' evaluations of natural resource use and wellbeing impacts of the REDD+ and NFCP intervention.

	(1) REDD+ Natural resource use impact	(2) NFCP Natural resource use impact	(3) REDD+, positive wellbeing impact	(4) NFCP, positive wellbeing impact
Land use variables				
Deforestation dummy	1.01 (1.17)	-0.73 (0.83)	-0.65 (0.86)	1.05 (0.80)
Agricultural land (log)	-1.17 (1.06)	0.12 (0.77)	-2.14* (1.26)	-0.67 (0.87)
Forest land (log)	0.33 (0.47)	-0.92 (0.85)	0.52 (0.71)	1.59** (0.79)
Income				
Total income (log)	-0.48 (0.62)	0.43 (0.62)	1.75*** (0.67)	0.50 (0.60)
Market access: Percent of cash income	-0.67 (1.39)	4.13*** (1.52)	0.54 (1.75)	0.83 (1.00)
Assets				
Total value assets (log)	-0.35 (0.28)	-0.95** (0.39)	-0.27 (0.35)	-1.27* (0.70)
Demographic variables				
Mean education of hh members	0.47* (0.29)	0.25 (0.25)	0.22 (0.20)	0.16 (0.14)
Mean age of hh members	-0.03 (0.04)	0.04 (0.05)	0.01 (0.05)	-0.07* (0.04)
Household size	0.03 (0.23)	0.16 (0.22)	-0.06 (0.25)	-0.05 (0.14)
Constant	4.01 (4.79)	0.82 (5.46)	-14.79** (5.81)	7.30 (6.85)
Observations	61	73	49	83
Village fixed effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.39	0.25	0.20	0.25
p-value	0.62	0.06	0.17	0.58

Table. 4.6. Self-reflexive evaluation of program impact on resources use and wellbeing (number of mentions is indicated in parenthesis)

REDD+		NFCP	
Impact on natural resource use	**Impact on wellbeing**	**Impact on natural resource use**	**Impact on wellbeing**
– Improved forest management (timber extraction, secondary forest's management, new knowledge on agroforestry practices) (21). – Better monitoring and taking care of forest (12). – Do not expand agricultural land on primary forest (11). – Do not cut old trees anymore (5). – Provision of equipment at the village level (chainsaw, boat, motor) (5). – Less hunting (1).	Negative – Did not comply with what was promised, no direct benefits to the households (9). – It is not clear how the money will be disbursed or about the budget (4). – Language is too technical, hard to understand (4). – Forest use is more limited (5). Positive – Community is better equipped to monitor forests (20). – Expecting disbursement of funds, income from sale of carbon credits (15). – Received capacity training, for improved forest management (7). – Community is equipped with transportation vehicle and chainsaw (6).	– Better conservation and management of forests (22). – Increased and/or improved ecotourism and fishing activities (22). – Improved monitoring of the rainforest (e.g., GPS, a drone) (21). – Improved forest management through reforestation and timber extraction (18). – We don't establish agricultural plots/ hunt (8) – Increased tenure security (3).	Negative – The benefits were not enough, and/or no direct benefits to the households yet (5). – There is no/little compensation for conducting monitoring activities (5). – Not everyone knows how to use the monitoring equipment (1). – Concern about the long-term continuity of the support and tenure security (1). Positive – Improved monitoring of forest (35) – Received capacity training and/or materials for artisanal products (30). – Received farm products (ducks-pigs) (19). – Received capacity training for reforestation projects, and expecting to extract timber from authorized area (17) – Delivery of items for fishing, capacity trainings (8)

4.5 Discussion

4.5.1 Program take-up and household participation in sustainable activities

Our results provide evidence that there is self-selection in collective agreements. This hypothesis is supported by the fact that a higher income from agricultural activities was associated with lower program participation. Furthermore, previous deforestation activities reduced the probability to participate in the NFCP. The fact that households who use more of the collective resources under the conservation agreement are less likely to participate in collective PES has been evidenced in other studies (Murtinho and Hayes, 2017; Hayes et al., 2019).

The negative correlation between program participation and conservation costs can hinder the effectiveness and the additionality of the program (Jack and Jayachandran, 2019), and brings forward the importance of improving within community targeting as a way to increase program effectiveness. Furthermore, and contrary to other studies (Kazungu et al., 2021), we find evidence that higher market access increases program participation. Finally, we note that even though we only considered economic variables as determinants of program participation, pro-social and pro-environmental preferences can also affect the decision to participate in conservation programs (Authelet et al., 2021).

4.5.2 Lack of impact: context, design, or implementation?

Overall, we find no evidence of anticipation effects, or negative behavioural changes in land use or income despite the delayed payments and slow project implementation. Taken together our results show that interventions can have positive effects on forest land use management but have limited impact on income and livelihood variables. A limitation of the study is that lack variables that measure the participants expectations of the program, and thus we cannot test or control for the mechanism in which the interventions are affecting livelihood and land use variables. This can also explain why we find no significant results. A second limitation is our relatively small sample size, which might be hindering finding significant effects. Notwithstanding, the results are consistent with previous evaluations indicating either none or small positive effects of collective agreements (Alix-Garcia et al., 2015; Naidoo et al., 2016). Mindful that the projects are still at an early phase, we highlight three sets of factors that explain why the incentive-based program had limited effects: contextual, design and implementation (Börner et al., 2017).

Regarding contextual factors, it is important to recall that the programs are implemented in communities and households with relatively low deforestation rates (~0.5 ha per year per household). The low *ex-ante* deforestation (i.e., a low baseline) and the subsistence nature of agricultural activities might explain the lack of strong impact. Households have few alternatives and shadow wages increase significantly when they are close to subsistence level (Angelsen 1999), and thus can expect limited changes in land use.

Moreover, the programs are implemented in contexts with overall low market integration, which also hinders the positive effects of investments.

In terms of design factors, the incentives provided by the program are simply too small to provide significant increases in income, as highlighted in the self-reflexive evaluations. In general, higher payments and compensations as well as lower transaction costs are, unsurprisingly, associated with a higher positive livelihood impacts of PES programs (Liu and Kontoleon, 2018; Ola et al., 2019). Even though cash payments - as compared to in-kind - can reduce the transaction cost of the program (Engel, 2016), and are often preferred by PES participants (Costedoat et al., 2016; Hayes et al., 2019), there are legal restrictions in Peru that constrain how the funds from the NFCP can be channelled to communities (Montoya-Zumaeta et al., 2021). Facing such legal constraints, ICDP-PES projects could alternatively explore possibilities to promote non-farm activities as the main mechanism to reduce the pressure on the natural resources, as this is also related to improved livelihood outcomes (Liu and Kontoleon, 2018).

Lastly, two main implementation factors delay household level impacts on land use and livelihoods. First, legal and administrative burdens imposed by the programs delay the distribution of benefits. For example, the administrative and legal requirements for the extraction of forest resources imposed by the NFCP, as well as the nature of the investments to implement the productive activities of the REDD+ program significantly increased implementation costs. The complexity of some collective agreements can in fact, threaten the enrolment and continued participation of communities and households (Izquierdo-Tort et al., 2021).

A second implementation challenge is that the benefits from the goods and assets provided at the village level (e.g., trucks, chainsaws), might not be trickling down evenly to all households. Evidence suggests that PES projects promoting communal infrastructure are less likely to provide positive livelihood benefits, or that they might take longer to materialize (Liu and Kontoleon, 2018). To understand why, examining within-community power relations is particularly important in collective agreements (Schröter et al., 2018), as elite capture can hinder the equitable distribution of benefits in collective agreements (Persha and Andersson, 2014; Almeida-Leñero et al., 2017). Good community governance and social capital (Krause et al., 2013), as well as reducing the burden of participation during program implementation, are thus likely to be essential components to maximize livelihood impacts.

4.5.3 The importance of participant's self-reflexive evaluations.

Incentive-based programs can provide broader ecological and well-being benefits than just income and land-use impacts (Hayes et al., 2019). Such non-monetary considerations include increased knowledge on ecosystem services (Arriagada et al., 2018) intrinsic motivations to conserve (Palmer et al., 2020), or improved perceived tenure security (Jones et al., 2020), or improved social capital (Alix-Garcia et al., 2018). We found evidence of non-monetary benefits associated with the program that are not captured by the monetary and land use indicators.

The perceived positive effects on wellbeing were related to increased monitoring capacities on their communal land, as well as more knowledge on the use and management of forest resources. Increased monitoring stands out as a particularly important and often mentioned component. This can be explained by the fact that indigenous communities in Peru are exposed to external illegal loggers, illegal agricultural settlements, as well as outside commercial fishers (Rodriguez-Ward and Paredes, 2014; Shanee and Shanee, 2016). External interventions that aim to increase community monitoring have shown to significantly contribute to better conservation of communal forests (Slough et al., 2021a), and we provide evidence that they also affect wellbeing. Both programs provided villagers with new technologies or more information to control and manage their forest resources. Similarly, providing information about best forest management practices, administrative and legal requirements, empowered villagers to manage their use of forest resources more sustainably in the short and medium term, while also increasing the amount of forest land used.

The reasons related to negative perceived impacts on wellbeing were mostly related to procedural aspects such as a lack of transparency and clarity during program implementation, or low overall compensation. The results corroborate previous evidence showing how low perceived levels of equity or unmet expectations can erode participants' wellbeing and program impact (Pascual et al., 2014; Montoya-Zumaeta et al., 2019), potentially threatening the continued implementation of the project. The evidence provided here highlights that ICDP-PES implementers should be careful of accurately communicating program implementation and provide consistent follow-up and consultations with participating villages and households, as this affects wellbeing. Adequate participation is in fact an important component of REDD+ safeguards (Duchelle et al., 2017). A closer collaboration between project implementers and participating households can mitigate some of these shortfalls (Schröter et al., 2018).

Overall, the results highlight the importance of considering household self-reflexive evaluations simultaneously to quasi-experimental evaluations. To date, most program evaluations in Latin America use only quantitative methods (Perevochtchikova et al., 2021) and focus on monetary livelihood outcomes, with there is less stress on non-monetary considerations (Blundo-Canto et al., 2018). Further, conservation research is skewed toward distributional concerns and to a lesser extent procedural dimensions (Friedman et al., 2018). We show that the personal experiences and procedural dimensions are important determinants of intervention impact on wellbeing.

4.6 Conclusion

This study investigates intra-community dynamics of participation and impact of collective agreements by evaluating two collective ICDP-PES projects in Ucayali, Peru, at early stages of implementation. It shows that there is within community self-selection into the program's activities, with households who are less dependent on agricultural activities more likely to participate in the program. It further evaluates whether there are anticipation effects on participating and non-participating households, finding that despite the slow and delayed implementation of the projects, there are no negative environmental impacts.

The results also show that the interventions are not affecting income, and we find suggestive evidence that they can have positive effects on the amount of households' forest land used. We highlight contextual, design and implementation factors that might hinder medium- and long-term impacts of both projects on income, and that are particularly relevant for collective agreements. Finally, the study brings forward the importance of considering household self-reflexive evaluations to both understand how an intervention might be having 'intangible' positive effects on wellbeing, such as increased monitoring of communal forest. Self-reflexive evaluations are also important to identify how procedural practices during program implementation affect household's evaluation of the intervention, and thus, potentially affect future and continued participation.

Acknowledgements

This research is part of CIFOR's Global Comparative Study on REDD+ (www.cifor.org/gcs). The funding partners that have supported this research include the Norwegian Agency for Development Cooperation (Norad, Grant No. QZA-16/0110 & No. QZA-21/0124), International Climate Initiative (IKI) of the German Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU, Grant No. 20_III_108), and the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA) with financial support from the CGIAR Fund Donors. We thank Javier Montoya, Santiago Izquierdo-Tort and KREM seminar participants for their valuable comments.

Appendix C

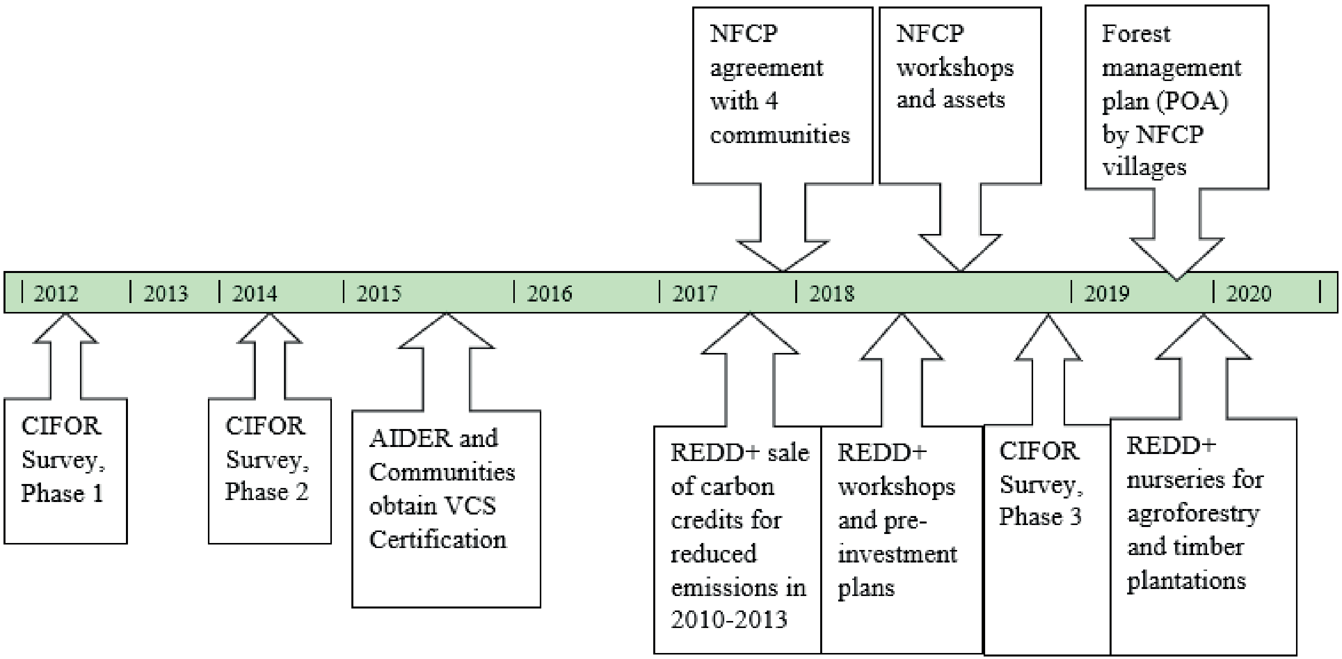
SECTION 1. Attrition analysis

Table 4.7. Attrition in the year 2012 and 2018

	(1) Attrition from 2014 to 2018	(2) Attrition from 2012 to 2014
Total income (log)	0.70* (0.14)	2.90** (1.35)
Cash income (% total income)	4.00*** (1.63)	0.85 (1.41)
Value of assets (log)	1.06 (0.15)	0.79 (0.15)
Forest land (has)	1.41 (0.35)	0.87 (0.22)
Agricultural land (has)	1.30 (0.24)	1.18 (0.45)
Mean education of hh members	1.09 (0.07)	1.42*** (0.10)
Mean age of hh members	0.88* (0.06)	0.87 (0.09)
Mean age of hh members (squared)	1.00** (0.00)	1.00 (0.00)
HH size	0.79* (0.11)	1.01 (0.20)
Household head gender	6.26** (4.62)	1.00 (.)
Village 1	3.02*** (0.63)	0.50** (0.17)
Village 2	0.63*** (0.08)	0.31*** (0.08)
Village 3	0.35*** (0.07)	1.64 (0.67)
Village 4	3.73*** (0.56)	0.80 (0.24)
Village 5	3.39*** (0.74)	4.22*** (2.32)
Village 6	1.56*** (0.22)	0.36*** (0.11)
Village 7	0.50*** (0.12)	0.59* (0.16)
Observations	244	181
Pseudo R squared	0.13	0.18

Note: Exponentiated coefficients; Clustered standard errors by village in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

SECTION 2. Timeline of the projects



SECTION 3. Matching results of for the average treatment effect of the treated (ATT).

We evaluated eight different matching algorithms using the command psmatch2 in Stata16 (Leuven and Sianesi, 2003): (i) 1 to 1 and 1 to 4 nearest neighbour with replacement, using the Mahalanobis distance, (ii) kernel propensity score matching using a normal, epanechnikov and biweight kernel, and (iii) radius matching with a caliper of 0.3 and 0.1, and (iv) local linear regression (LLR) matching. Compared to caliper or nearest neighbour matching, kernel matching and LLR matching allow to use the whole sample of control observations to construct the weighted average of the counterfactual match for each participant. All matching algorithms reduced Rubin's B and Rubin's R compared to the unmatched sample, but in general LLR performed best in the case of the REDD+ project and caliper matching performed best in the case of the NFCP.

a) ATT matching of the REDD+ participating households

Table 4.8. indicates the summary statistics of the matching methods evaluated for the REDD+ intervention and control observations. Before matching, the median bias is 14.9 and the mean bias is 11.2. The best matching strategy is obtained with the LLR matching, as it has the lowest Rubins B and a Rubins R lower than 2.

Table 4.8. Summary statistics of matching methods and results for the ATT of the REDD+ treatment.

Matching method REDD	Mean Bias	Median Bias	Rubins B	Rubins R	%Var
Unmatched	14.9	11.2	80.0	1.03	44
Nearest Neighbour					
M1NN	13.8	13.2	99.4	1.75	38
M4NN	15.6	15.9	100.2	1.20	31
Kernel					
Normal	6.2	6.5	29.5	1.29	25
Epenchikov	5.8	4.4	29.9	1.43	19
Biweight	6.0	5.5	28.7	1.48	19
Radius					
Caliper(0.1)	6.4	6.8	30.4	1.28	25
Caliper(0.3)	8.4	5.3	52.7	0.71	25
LLR matching	3.5	2.9	20.9	1.05	13

Fig. 4.3. indicates the standardized percentage bias of each covariate, before and after matching, and is used as a main indicator of covariate imbalance between the treatment and control observations. It is calculated (Rosenbaum and Rubin, 1985):

$$SD = \frac{\bar{X}_T - \bar{X}_c}{\sqrt{\frac{(S_T^2 + S_C^2)}{2}}}.$$

Where SD is the standardised difference in means, T and C represent treated and control groups, $\bar{X}$ their means, and S their standard deviations. The indicator has a value between 0 and 1, with one indicating greater imbalance. The variable with the highest standardized difference is the extent of participation in previous programs. However, matching satisfactorily corrects for the difference between treatment and control observations, reducing the standardized difference to below the 10% level for all covariates. Figure indicates the propensity score distribution of the observations. Overall, four treated observations are left out of analysis because they do not satisfy the common support assumption.

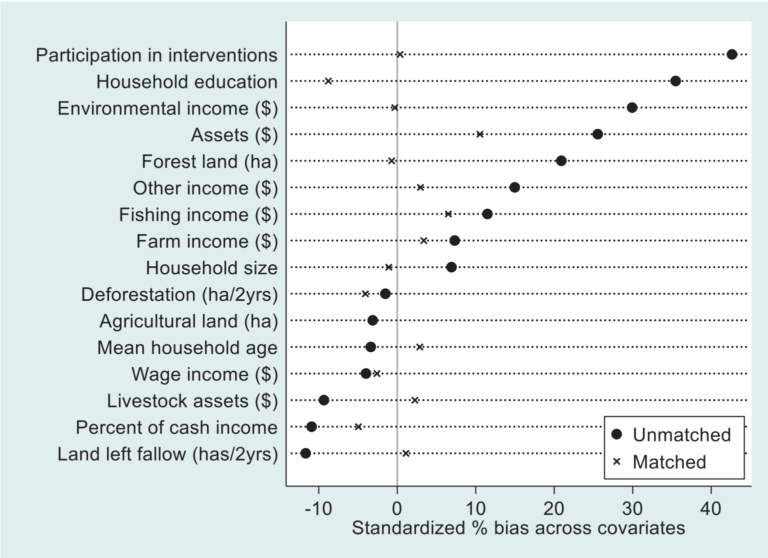


Figure 4.3. Standardized percent bias of covariates between the REDD+ treated and control observations.

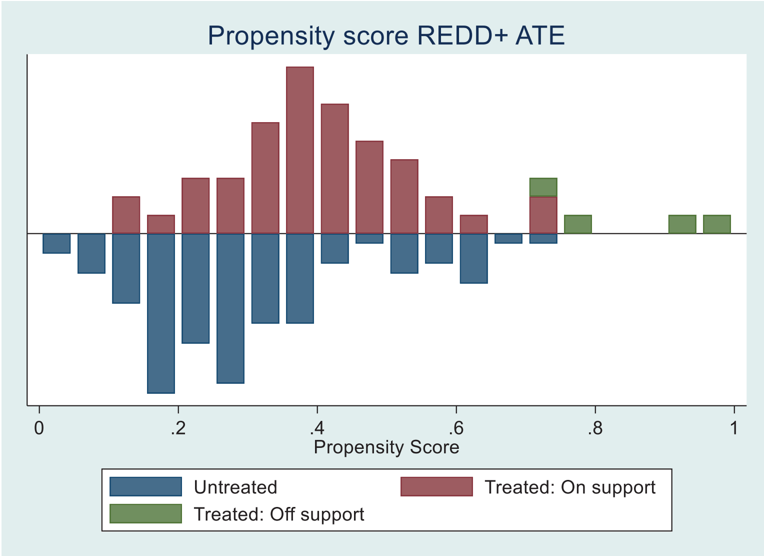


Figure 4.4. Common support between treatment and control observations considered for REDD+'s ATT.

b) ATT matching of the NFCP participant households

Table 4.9 indicates the summary statistics of the matching methods evaluated for the average treatment effect on the treated (ATT) of the NFCP intervention. Before matching, the mean bias is 36.2 and the median bias is 35.6 units. The best matching strategy is obtained with caliper matching, as it has the lowest Rubins B and a Rubins R, as well as the lowest mean and median bias. The covariate with the highest standardized percent bias is farm income (Figure), equivalent to -82.2% before matching and to --29.1% after matching. None of the variables have significant differences in means after matching, and all the variables (except farm income) have a standardized bias lower that 25%. Overall, there is less common support that in the REDD+ project (resulting in dropping 13 observations with no common support) between treated and control observations (Figure).

Table 4.9 Summary statistics of matching methods and results for the ATT of the NFCP treatment.

Matching method NFCP ATT	**Mean Bias**	**Median Bias**	**Rubins B**	**Rubins R**	**%Var**
Unmatched	36.2	35.6	164.1	0.69	38
Nearest Neighbour					
M1NN	10.4	8.7	118.5	1.76	56
M4NN	11.4	7.3	122.8	1.06	50
Kernel					
Normal	10.2	10.2	57.1	1.13	31
Epenchikov	12.9	9.8	72.5	0.75	31
Biweight	12.8	9.3	74.2	0.73	31
Radius					
Caliper(0.1)	13.0	9.6	69.3	0.63	31
Caliper(0.3)	9.9	7.7	60.2	0.92	31
LLR matching	13.7	10.8	74.6	0.72	31

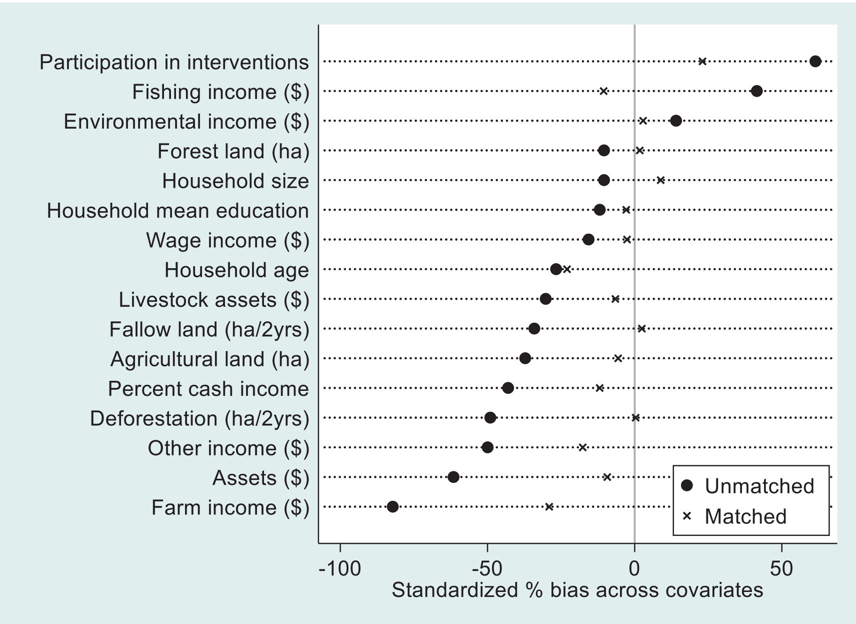


Figure 4.5. Standardized percent bias of covariates between the NFCP treated and control observations.

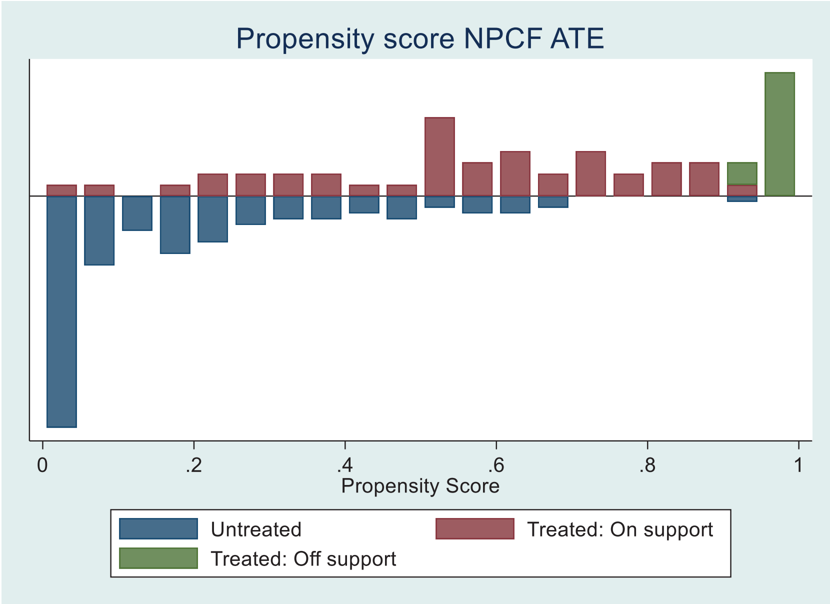


Figure 4.6. Common support between treatment and control observations considered for the NFCP's ATT.

SECTION 4. Matching results of non-participating households.

a) Matching of the REDD+ non-participating households.

We conducted a separate matching analysis for the non-participating households in the REDD+ and NFCP intervention villages. The best reduction in standardized bias were obtained with LLR matching, which reduced Rubin's B to 27 in case of the REDD+ intervention and to 24.5 in case of the NFCP intervention.

Table 4.10. Summary statistics of matching methods and results for the ATNP of the REDD+ treatment.

Matching method REDD+ ATNP	Mean Bias	Median Bias	Rubins B	Rubins R	%Var
Unmatched	19.0	14.4	84.6	1.5	38
Nearest Neighbour					
M1NN	13.8	12.7	74.5	1.93	38
M4NN	16.1	13.2	77.1	1.31	44
Kernel					
Normal	5.2	3.3	26.9	0.73	13
Epenchikov	5.7	3.8	29.5	0.62	13
Biweight	6.1	4.2	31.1	0.58	13
Radius					
Caliper(0.1)	6.5	4.0	33.7	0.53	13
Caliper(0.3)	9.6	7.8	45.3	1.65	13
LLR matching	4.6	5.0	27.0	0.72	13

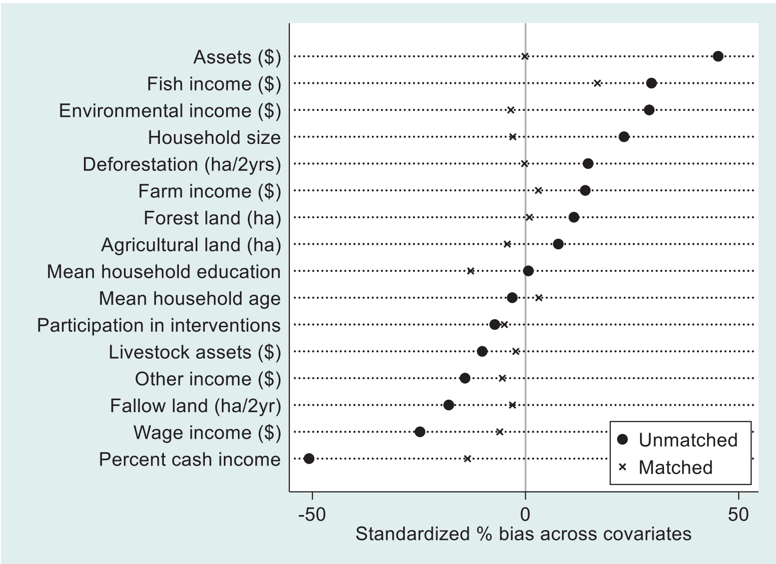


Figure 4.7. Standardized percent bias of covariates between the REDD non-participating and control observations.

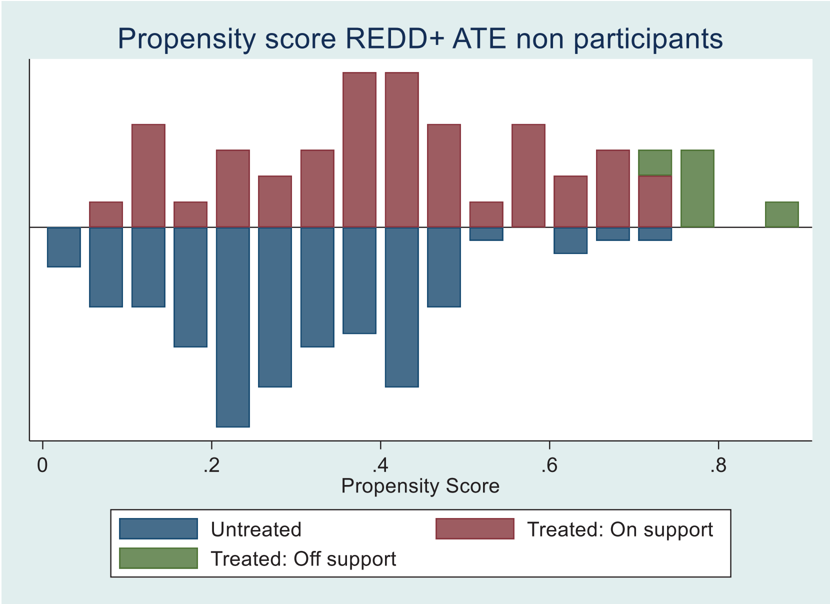


Figure 4.8. Common support between treatment and control observations considered for REDD+'s ATNP.

b) Matching of the NFCP non-participating households.

Table 4.11. Summary statistics of matching methods and results for the ATNP of the NFCP treatment.

Matching method NFCP ATNP	**Mean Bias**	**Median Bias**	**Rubins B**	**Rubins R**	**%Var**
Unmatched	200.8	21.0	109.7	0.27	25
Nearest Neighbour					
M1NN	12.7	10.3	107.0	1.03	19
M4NN	17.8	20.0	96.5	0.69	25
Kernel					
Normal	9.4	6.8	56.5	0.67	19
Epenchikov	9.3	6.2	54.9	0.71	19
Biweight	9.2	6.4	55.6	0.76	19
Radius					
Caliper(0.1)	9.2	5.3	54.2	0.81	19
Caliper(0.3)	15.6	14.9	81.2	0.18	19
LLR matching	5.7	4.2	24.5	2.34	13

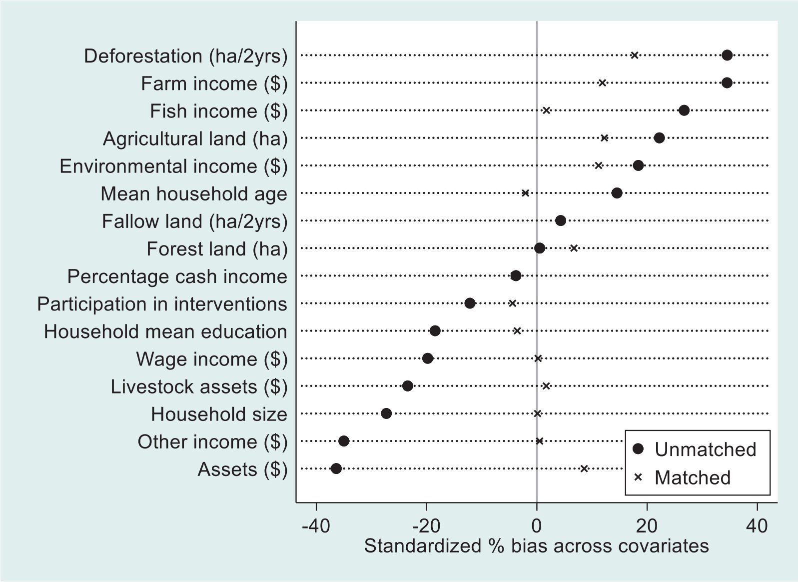


Figure 4.9. Standardized percent bias of covariates between the NFCP non participating and control observations.

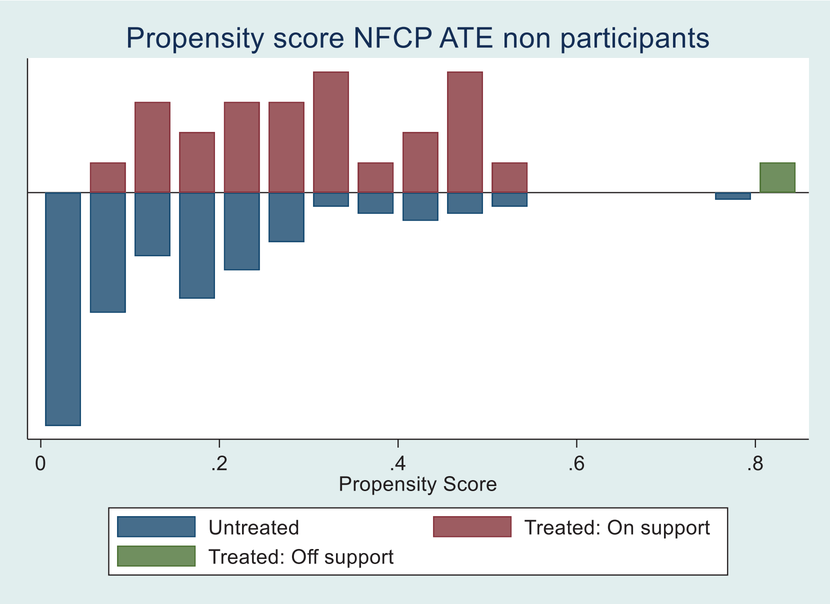


Figure 4.10. Common support between treatment and control observations considered for the NFCP's ATNP.

SECTION 5. Matching results of the Intention to Treat (ITT)

a) Intention to treat (ITT) matching of the REDD+ intervention.

The matching results indicate that the best matching method is LLR matching, reducing the mean bias to 4.0 and median bias to 2.8, and Rubins B to 21.6 and Rubins R to 1.06. Out of the total sample of 187, six treated observations are dropped because they are off common support.

Table 4.12. Summary statistics of matching methods and results for the ITT of the REDD+ treatment.

Matching method REDD+ ITT	Mean Bias	Median Bias	Rubins B	Rubins R	%Var
Unmatched	15.4	14.8	70.6	1.35	50
Nearest Neighbour					
M1NN	11.3	11.6	71.3	1.60	56
M4NN	13.0	11.4	77.1	1.35	63
Kernel					
Normal	5.2	4.5	24.6	1.31	19
Epenchikov	5.4	5.2	25.9	1.58	25
Biweight	5.3	5.3	24.8	1.47	25
Radius					
Caliper(0.1)	5.3	4.6	24.8	1.26	19
Caliper(0.3)	7.8	7.2	45.0	1.26	25
LLR matching	4.0	2.8	21.6	1.06	25

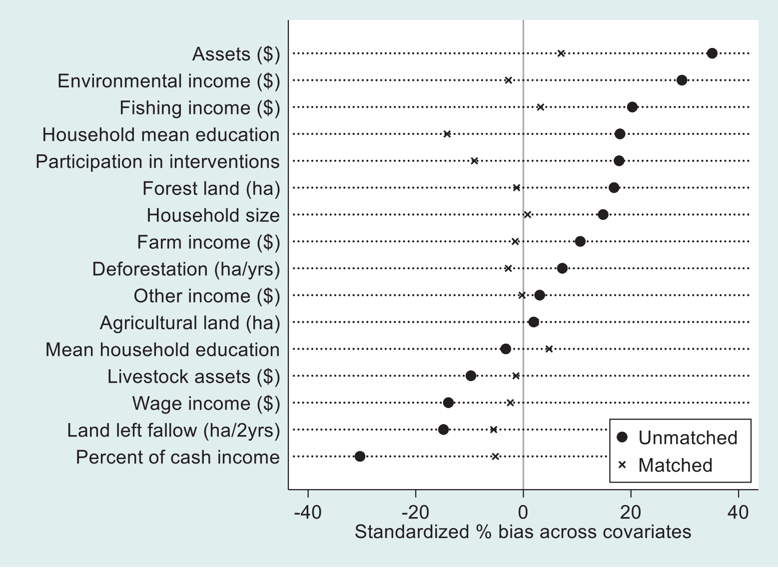


Figure 4.11. Standardized percent bias of covariates between the REDD village households and control observations.

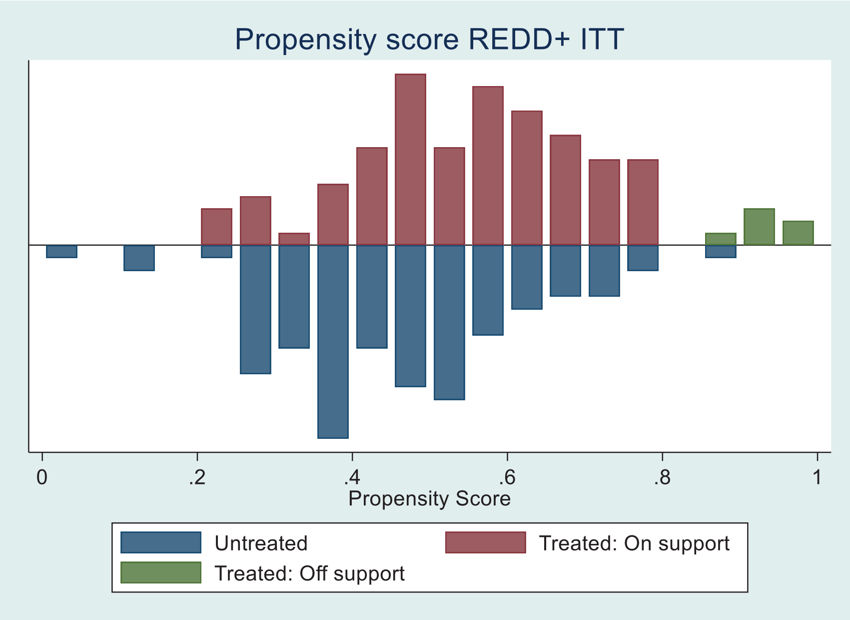


Figure 4.12. Common support between treatment and control observations considered for REDD+'s ITT.

b) Intention to treat (ITT) of the NFCP interventions

The unmatched sample of the NFCP ITT sample has a mean vias of 24.9 and a median bias of 18.7. The matching results of the intention to treat sample yielded better outcomes than the matching of the average treatment effect. The caliper matching with 0.1 as radius perform best, reducing Rubins B to 27.7, with an R of 1.65. A total of 16 observations are dropped because no common support.

Table 4.13. Summary statistics of matching methods and results for the ITT of the NFCP treatment.

Matching method NFCP ITT	**Mean Bias**	**Median Bias**	**Rubins B**	**Rubins R**	**%Var**
Unmatched	24.9	18.7	110.1	0.72	50
Nearest Neighbour					
M1NN	8.2	8.3	68.0	1.34	38
M4NN	9.4	7.9	81.3	0.74	63
Kernel					
Normal	6.7	5.3	28.3	1.81	50
Epenchikov	6.7	5.6	28.4	1.70	44
Biweight	6.6	5.4	28.5	1.64	44
Radius					
Caliper(0.1)	6.6	6.0	27.7	1.65	44
Caliper(0.3)	8.0	7.0	41.6	0.71	38
LLR matching	7.0	6.7	29.0	1.54	44

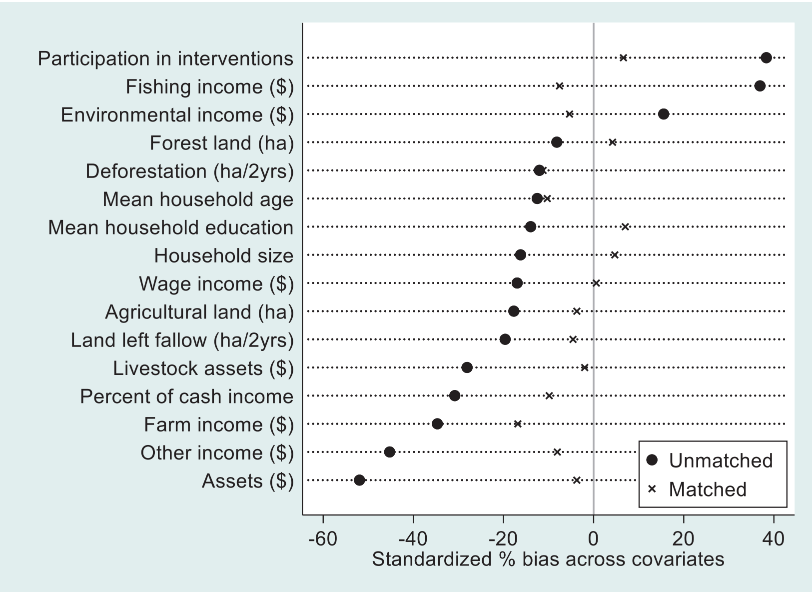


Figure 4.13. Standardized percent bias of covariates between the NFCP village households and control observations.

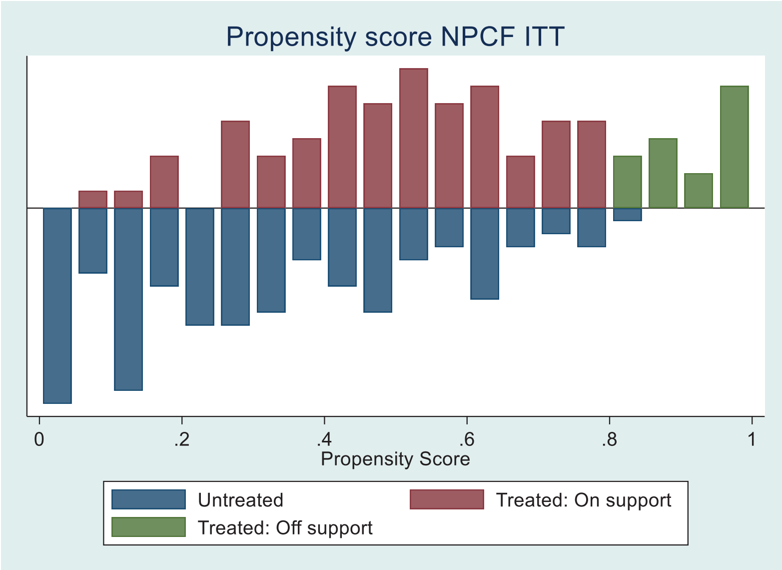


Figure 4.14. Common support between treatment and control observations considered for NFCP's ITT.

SECTION 6. Tests of parallel trends

Table 4.14. Test for parallel trends assumption of REDD+ ATT control and intervention observations

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
REDD+ ATT	-0.16	0.06	-0.45	0.09	109.79	-972.93	308.94
	(0.49)	(0.35)	(0.34)	(0.37)	(622.14)	(1189.49)	(280.34)
NFCP (control)	0.32	0.30	0.70*	-0.03	577.06	1721.52	-244.80
	(0.26)	(0.23)	(0.31)	(0.18)	(407.04)	(1162.86)	(354.88)
Constant	-0.59	-0.32	-0.56*	-0.26	-1063.46	-2151.34**	58.13
	(0.33)	(0.17)	(0.27)	(0.37)	(592.03)	(734.18)	(283.12)
Observations	101	105	104	107	103	102	104
Adjusted R^2	-0.011	-0.015	0.012	-0.018	-0.007	0.006	0.004
p-value	0.50	0.12	0.12	0.94	0.40	0.35	0.50

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.15 Test for parallel trends assumption of NFCP ATT control and intervention observations.

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
NFCP ATT	-0.17	-0.58	0.39*	-0.28	164.14	1747.35	-347.70
	(0.34)	(0.49)	(0.18)	(0.22)	(400.44)	(1994.27)	(299.94)
REDD+ (control)	-0.01	-0.43	-0.89*	0.25	-299.21	-415.76	612.02
	(0.26)	(0.54)	(0.39)	(0.13)	(691.44)	(2131.00)	(362.78)
Constant	-0.08	0.49	-0.28*	-0.13	-834.32**	-4209.92***	-160.47
	(0.23)	(0.45)	(0.13)	(0.09)	(251.28)	(1081.17)	(310.11)
Observations	116	119	121	118	116	115	117
Adjusted R^2	-0.012	0.015	0.064	0.035	-0.008	-0.014	0.027
p-value	0.88	0.47	0.09	0.25	0.86	0.63	0.03

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.16 Test for parallel trends assumption of REDD+ ATNP control and interventions observations.

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
REDD+ ATNP	0.49	0.99^{*}	0.29	0.09	602.15	-385.61	-649.61
	(0.28)	(0.42)	(0.39)	(0.27)	(427.18)	(2356.71)	(403.18)
NFCP participation (control)	0.20	-0.35	0.51	-0.53^{**}	238.39	440.39	246.02
	(0.24)	(0.37)	(0.42)	(0.18)	(271.76)	(2854.75)	(327.61)
Constant	-0.53^{*}	-0.23	-0.47	0.14	-960.00^{**}	-3915.53^{*}	-210.33
	(0.23)	(0.23)	(0.43)	(0.21)	(359.30)	(1753.47)	(241.33)
Observations	102	105	105	103	101	102	102
Adjusted R^2	0.021	0.021	0.013	0.032	0.015	-0.020	0.004
p	0.21	0.13	0.50	0.04	0.22	0.88	0.31

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.17. Test for parallel trends assumption of NFCP ATNP control and intervention observations.

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
NFCP ATNP	-0.93	-0.90^{*}	-0.89^{**}	-0.24	-851.17	-915.27	-840.59
	(0.58)	(0.47)	(0.26)	(0.57)	(857.27)	(943.57)	(561.19)
REDD+ participation (control)	-0.98	-1.46^{**}	-1.20	0.40	-1754.12	2146.13^{*}	693.39
	(1.02)	(0.56)	(0.66)	(0.36)	(1005.98)	(1117.35)	(459.27)
Constant	0.09	0.86^{*}	0.11	0.10	-41.46	-3776.67^{***}	-108.74
	(0.30)	(0.43)	(0.12)	(0.14)	(321.20)	(549.51)	(299.69)
Observations	103	104	105	109	102	102	103
Adjusted R^2	0.098	0.062	0.030	-0.000	0.063	-0.017	0.035
p	0.05	0.05	0.03	0.46	0.00	0.20	0.31

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.18. Test for parallel trends assumption of REDD+ ITT control and interventions observations.

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
REDD+ ITT	0.29	0.70	-0.16	0.15	386.05	-1773.99	-380.77
	(0.27)	(0.44)	(0.29)	(0.31)	(443.45)	(1130.09)	(326.23)
NPCF villages	-0.39	-0.11	0.10	-0.50	-397.22	835.42	-147.58
	(0.31)	(0.43)	(0.28)	(0.33)	(457.90)	(1247.27)	(324.08)
Constant	-0.36**	-0.40	-0.36	-0.01	-731.17*	-2148.41*	38.97
	(0.14)	(0.38)	(0.34)	(0.21)	(357.29)	(911.17)	(267.55)
Observations	141	143	144	145	140	139	140
Adjusted R^2	0.007	0.011	-0.012	0.034	0.002	0.007	-0.001
p	0.27	0.33	0.08	0.36	0.63	0.03	0.42

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.19 Test for parallel trends assumption of NFCP ITT control and interventions observations.

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
NFCP ITT	-0.57	-0.98**	-0.24	-0.52	-465.45	2349.65	-538.81
	(0.39)	(0.33)	(0.25)	(0.28)	(434.97)	(1638.37)	(293.21)
REDD+ villages	0.26	0.43	-0.49*	0.31	-15.49	-259.78	62.85
	(0.39)	(0.33)	(0.25)	(0.29)	(452.43)	(1608.05)	(294.54)
Constant	-0.19	0.42*	-0.03	-0.03	-551.15	-5012.81**	20.39
	(0.27)	(0.20)	(0.14)	(0.23)	(353.05)	(1891.20)	(218.99)
Observations	134	135	137	138	134	132	134
Adjusted R^2	0.022	0.029	0.001	0.069	-0.005	-0.012	0.009
p	0.32	0.04	0.11	0.25	0.48	0.24	0.22

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

SECTION 6. Main results

Table 4.20. Average treatment effect on the treated (ATT) of REDD+

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
REDD+ participant	0.16	0.13	-0.07	0.14	86.51	229.62	-239.58
	(0.12)	(0.29)	(0.23)	(0.16)	(86.17)	(207.53)	(149.83)
NFCP participation (as control)	-0.08	1.02^{***}	0.36^{*}	0.02	356.97^{*}	-142.42	245.99^{*}
	(0.09)	(0.27)	(0.17)	(0.10)	(178.07)	(324.66)	(123.43)
Constant	-0.14	-1.54^{***}	-0.30^{***}	-0.01	-308.07^{***}	-564.41^{**}	-42.33
	(0.08)	(0.25)	(0.06)	(0.14)	(76.79)	(191.63)	(46.22)
Observations	124	126	124	127	124	124	124
Adjusted R^2	-0.008	0.052	0.025	-0.006	0.075	0.002	0.010
p	0.46	0.02	0.03	0.71	0.02	0.13	0.21

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.21. Average treatment effect on the treated (ATT) of the NFCP, with regression adjustment

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
NFCP participant	-0.43	0.82^{*}	-0.31^{*}	-0.03	53.88	237.70	102.88
	(0.36)	(0.36)	(0.14)	(0.16)	(117.67)	(220.20)	(151.04)
REDD+ participation (as control)	-0.10	-1.00^{**}	-0.12	-0.16	224.88	-92.24	-483.17^{*}
	(0.38)	(0.35)	(0.23)	(0.11)	(119.11)	(263.83)	(229.61)
Constant	0.21	-1.64	0.16	0.04	159.79	632.70	86.50
	(0.37)	(0.87)	(0.34)	(0.19)	(163.05)	(338.01)	(191.41)
Observations	135	136	139	136	135	135	135
Adjusted R^2	0.047	0.132	0.107	0.017	0.291	0.317	0.268
p-value	0.05	0.13	0.11	0.02	0.29	0.32	0.27

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.22 Average treatment effect on the non-participating (ATNP) of REDD+, with regression adjustment

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
REDD+ non participant	0.35 (0.27)	0.62** (0.26)	0.27 (0.21)	0.22 (0.13)	159.26 (99.19)	78.70 (184.17)	72.52 (188.35)
NFCP participation (as control)	-0.06 (0.21)	1.21*** (0.30)	0.15 (0.29)	0.24* (0.12)	478.71*** (94.33)	362.49** (137.01)	105.61 (152.58)
Constant	-0.56* (0.28)	-1.93** (0.66)	-0.85** (0.27)	-0.58 (0.40)	-544.20* (265.26)	388.11* (186.22)	326.88 (277.63)
Observations	121	124	124	123	121	121	121
Adjusted R^2	-0.008	0.065	0.013	0.037	0.089	0.617	-0.015
p	0.16	0.00	0.01	0.18	0.01	0.00	0.02

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.23. Average treatment effect on the non-participating (ATNP) of the NFCP, with regression adjustment

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
NFCP non-participant	0.33 (0.25)	0.64 (0.40)	0.03 (0.17)	-0.09 (0.15)	-24.98 (121.24)	-48.95 (350.05)	-36.33 (238.44)
REDD+ participation	-0.07 (0.33)	0.53 (0.29)	-0.33 (0.19)	0.04 (0.09)	-119.44 (148.32)	235.15 (421.72)	-455.24*** (117.29)
Constant	-0.21 (0.31)	-1.20** (0.43)	0.03 (0.20)	0.47*** (0.11)	133.30 (87.42)	-289.86 (317.51)	-123.09 (227.34)
Observations	119	120	121	124	119	119	119
Adjusted R^2	0.284	0.090	0.095	0.670	0.346	0.155	0.164
p	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.24 Intention to Treat (ITT) of REDD+

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
REDD+ ITT	0.11	0.17	0.08	0.20	57.54	8.78	18.20
	(0.14)	(0.19)	(0.10)	(0.12)	(104.48)	(102.68)	(107.41)
NPCF villages	0.01	0.84***	0.17	0.12	197.37	-193.09*	16.96
	(0.16)	(0.17)	(0.10)	(0.12)	(109.55)	(97.26)	(108.34)
Constant	-0.22	-1.59***	-0.29	-0.27	-387.17**	-710.10*	244.31
	(0.25)	(0.35)	(0.18)	(0.20)	(149.11)	(305.69)	(260.84)
Observations	164	165	167	167	163	163	163
Adjusted R^2	-0.015	0.050	-0.005	0.018	0.017	-0.004	-0.014
p	0.63	0.00	0.20	0.33	0.36	0.01	0.57

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.25 Intention to Treat (ITT) of NFCP

	(1) Ag. Land (ha)	(2) Forest land (ha)	(3) Land left fallow (ha in last 2 years)	(4) Reported forest clearing (ha in last 2 years)	(5) Farm income (ha)	(6) Environmental income (ha)	(7) Other income (ha)
NFCP ITT=1	0.02	0.61**	-0.20	-0.12	25.70	112.74	222.25
	(0.30)	(0.24)	(0.11)	(0.13)	(105.31)	(124.87)	(206.28)
REDD+ villages	0.18	0.17	0.10	0.19	20.41	-173.29	-264.24
	(0.27)	(0.21)	(0.10)	(0.11)	(124.41)	(164.76)	(203.24)
Constant	-0.40	-1.08**	0.20	0.18	115.12	-382.28**	-174.98
	(0.33)	(0.37)	(0.15)	(0.16)	(93.52)	(131.87)	(169.23)
Observations	155	156	158	158	155	155	155
Adjusted R^2	0.011	0.069	0.104	0.329	0.242	0.010	0.073
p	0.42	0.00	0.00	0.00	0.00	0.00	0.09

Note: Clustered standard errors at the village level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Village focus group discussion	**REDD+**	**NPCF**
Impact on natural resource use	No effect (2 villages) Positive effect (2 villages): Better monitoring equipment, less insertion of illegal loggers. They bought equipment to cut timber and produce sawn wood.	Positive (4 villages): They are now conserving and monitoring their forests; They are starting to undertake reforestation activities with species "bolaina", a fast-growing timber species. They are better conserving and monitoring their forests. With the money from the program, they bought a drone and received capacity training from Rainforest Foundation. There is income from fishing activities, provided nets, they are also working on a plan to extract timber at a small scale, and they have also built a guest house for tourists. They would like, however, more time and resources dedicated to training and capacity building. They think the monitoring is good because impedes illegal settlers however they would like to receive compensation, they only obtain food to conduct the monitoring activity.
Impact on wellbeing	Negative (1 village): “There wasn’t participation of the population on the total budget of the project, everything was decided by NGO’s central office”. Lack of clarity and feelings of exclusion. Positive (1 village): They can improve the construction material for their houses, and they are also expecting payments. Neutral (2 villages): There is too much information in the workshops, it is too technical. They received tools for land demarcation such as GPS, cameras, but they don’t know how much money is being received and imbursed/invested in the community. They would rather receive the money directly and have no intermediaries.	Positive (4 villages): They started to farm pigs and they received materials and inputs to produce artisan crafts. They are improving their use if the forest, can now extract timber, and have the resources and capital to produce sawn wood. They havent sold anything yet but they hope they will be able to do so in the future. The other activities, such as ecotourism, are still to be implemented. They also have more food security since families received ducks. Although it is good, they received pigs, they did not provide food for the pig, didn’t provide technical assistance. They gave received higher incomes from selling artisan products, and they hope that in the future ecotourism will also bring more income. It has allowed to obtain transportation vehicle, and capital and chainsaws to transform timber.

5 Paper IV

"We abstract heroically in more ways than one". (M. Weitzman, 1976)

Re-examining the macroeconomic drivers of deforestation

Julia Naime[1,2]
junaime@nmbu.no

[1]School of Economics and Business, Norwegian University of Life Sciences (NMBU), Ås, Norway
[2]Center for International Forestry Research (CIFOR), Bogor, Indonesia

Abstract

Drawing from dual economy growth theories, this paper presents an empirical approach and theoretical consistent framework to analyse the macroeconomic determinants of deforestation. This represents a novel approach to study the drivers of deforestation, in two ways. First, it examines the effect of structural change - defined as changes in the relative returns of factors of production across a 'modern' and a 'traditional' sector - on deforestation rates. Second, it distinguishes the direct drivers, which are the changes in the relative returns of factors of production across sectors, from the underlying drivers of deforestation, which are the factors that determine the changes in the relative returns of factors of production across sectors. The empirical strategy consists of three steps. First, I estimate marginal returns of each factor of production (land, labour and capital) at the national level. Second, to identify immediate drivers of deforestation I examine the relationship between deforestation and changes in the marginal returns to land, labour and capital across sectors. Finally, I focus on trade as an underlying driver and evaluate associations between international trade variables (i.e., openness to trade and exchange rates) to changes in the relative marginal returns to each factor of production. The relative return between agricultural land and forest land is the single and most important immediate driver of deforestation. Openness to trade can, indirectly, reduce deforestation levels by decreasing the relative returns between agricultural land and forest land.

Keywords: deforestation, growth, structural change, agriculture, forest
JEL codes: C23, Q23, Q24, F18, O44

5.1 Introduction

Forests cover 30.8% of the land area of our planet (FAO, 2020). Forests and their biodiversity provide essential ecosystem services that contribute to our global economy, including water regulation, carbon capture and storage (Harris et al., 2021), and provision of raw materials such as food and timber (Angelsen et al., 2014). Forest use and its conversion have been a source of growth for many countries, but in a world with climate change, their continued loss has become a major concern. Deforestation and forest degradation have taken centre stage in the climate mitigation debate, as continued forest loss both contributes to emissions and also threatens the carbon sink potential of forests (Mitchard, 2018). While East Asian tropical forests have become a net source of emissions while the Amazon and Congo basin are still net carbon sinks (Harris et al., 2021).

The estimated annual deforestation was 10 million ha for the years 2015-2020 (FAO, 2020). Deforestation is mostly concentrated in the tropics, and the pattern of forest loss and the drivers of deforestation vary across countries and over time. In the last decade, Africa had the highest deforestation, with 3.9 million ha per year, followed by South America with 2.6 million ha per year (FAO 2020). While deforestation rates in Africa are on the rise, they are decreasing in some areas of South America. On the other hand, other regions are experiencing net forest regrowth, such as Asia, Europe and Oceania (FAO 2020). Developing countries such as China, Vietnam, Nepal and Costa Rica have experienced net forest regrowth in the last decade (Meyfroidt et al., 2010)

Understanding why patterns of deforestation differ across regions and countries and identifying the drivers of land use change is key for the design of effective and efficient mitigation policies nationally and globally. Two major and related 'theories' – or rather empirical patterns - to explain long terms changes in forest cover are the Forest Transition and the Environmental Kuznets Curve of deforestation (Mather, 1992; Koop and Tole, 1999). They suggest that over time, as countries grow, deforestation increases but at a decreasing rate, and can eventually recover (Caravaggio, 2020a). National forest cover can thus follow a U-shaped pattern: areas that have high forest cover tend to experience high deforestation rates at early stages of economic development, while as forest cover decreases, deforestation rates decrease, and forest cover eventually increases again. The rate of forest conversion is ultimately determined by changes in the returns to competing land uses (Barbier et al., 2017).

Multiple cross-country comparisons have tested these theories, sometimes finding contradictory results (Caravaggio, 2020b). Economic growth is often hailed as one of the primary drivers of deforestation, along with population growth, trade, and institutional factors (Barbier and Tesfaw, 2015; Wolfersberger et al., 2015; Leblois et al., 2017). While these studies provide useful insights, a persistent challenge in the literature linking economic growth with deforestation has been to distinguish the direct drivers from the underlying drivers (Kaimowitz and Angelsen, 1998), as well as finding causal attribution and describing the mechanisms between economic growth and forest cover loss.

This paper is a contribution to the large literature that examines the drivers of deforestation at the national level. I apply a novel empirical approach to distinguish between the immediate drivers of deforestation and the underlying drivers, drawing from dual economy growth models. Section 5.2 presents the definition of drivers of deforestation and the theoretical framework with which I analyse deforestation at the macroeconomic level and disentangle the *immediate* drivers from the *underlying* drivers of deforestation. In section 5.3, the empirical strategy to distinguish the two types of drivers is presented, followed by the results and discussion in section 5.4 and 5.5 respectively.

5.2 Theoretical background

5.2.1 Definition of drivers of deforestation

I adopt the classification of drivers of deforestation developed by Angelsen and Kaimowitz (1999) that distinguishes a threefold hierarchy of causes of deforestation. The first is the *source* of deforestation, which refers to the land use activities that replace the forested area, such as agriculture (cropland and pasture), mining, or urban development. These are often referred to as the direct drivers of deforestation (Geist and Lambin, 2002). The second are the *immediate drivers* of deforestation, i.e., the decision parameters of the agents of deforestation. The immediate drivers can be summarized as the opportunity cost of conservation, which are determined by the relative value of forestland as compared to the highest alternative land use value, normally agricultural production. Finally, the *underlying drivers* of deforestation are the macroeconomic characteristics that affect the *immediate drivers*. The underlying drivers are broadly defined and can include economic and non-economic characteristics that affect the opportunity costs of forest land, such as institutions, trade policies, and infrastructure development.

Significant progress has been made over the past decades to uncover the *sources* of deforestation. Better remote sensing and satellite technologies have significantly improved our understanding of the sources of deforestation, by providing spatially detailed information of changes in land use at the national and subnational levels. Recent analysis of global satellite data distinguish among five different sources of forest cover loss (Curtis et al., 2018): commodity-driven agriculture, shifting agriculture, forestry, wildfires, and urbanization. Except for urbanization, each source of deforestation bears a roughly similar responsibility for total forest cover loss, of around ¼ each. A few studies have analysed the sources of deforestation on a regional scale. In South America, pasture has been the dominant source of deforestation, followed by commercial cropland (Sy et al., 2015). Analysis has also shown that the area and extent of deforestation events has increased, which suggests the increasing prominence of industrial scale driven deforestation (Austin et al., 2017).

Hence, over the years, the quality of economic and biophysical data to conduct cross country regressions has significantly improved, including the availability of time series data of deforestation (Hansen et al., 2013). These innovations have paved the road for improving econometric analysis of national level deforestation rates. Recent analyses of macroeconomic drivers showed that 'usual' drivers of deforestation such as population density, economic development and agricultural activity, explain deforestation rates in the 2000-2010 decade (Crespo Cuaresma et al., 2017; Leblois et al., 2017).

These economic analyses have, however, stagnated in terms of improving the understanding and empirical identification of the causal mechanism between growth and deforestation. In an early review of economic models of deforestation, Kaimowitz and Angelsen (1998) concluded that most cross-country analysis of drivers of deforestation suffered from several weaknesses. One of them was data quality, but another one was a mixing of direct and indirect causes of deforestation, in part reflecting the lack of an explicit theoretical framework. Indeed, most analyses bundle together the underlying and the immediate drivers using aggregate or agricultural GDP as the main independent variable, including control variables such as institutional variables, or trade openness.

The lack of an appropriate theoretical framework to conduct cross-country regressions has been pervasive in the empirical literature, resulting in the persistent overlay between immediate and underlying drivers of deforestation. The lack of theory hampers the capacity to identify the various channels through which growth affects forest cover change, and thus the macroeconomic policies that could help reduce deforestation. This can also explain why there are inconsistencies in understanding how economic growth affects deforestation rates, with some finding positive relationship and other negative ones (Culas, 2012). While one should not expect to find a straightforward relationship, we need to better understand why deforestation differs across an heterogenous tropical world.

5.2.2 Dual economies

This section presents the theoretical framework used for the empirical analysis. It describes the main impact mechanism by which economic growth affects deforestation rates. In particular, it investigates how structural changes in the economy can affect deforestation rates. Structural transformation is understood as the process by which factors of production (land, labour or capital), are reallocated into different activities over time and as economies grow. For the purposes of this analysis, the *source* of deforestation is restricted to agricultural expansion, which is the most relevant source of deforestation globally (Hosonuma et al., 2012; Curtis et al., 2018). For simplicity, the theoretical model considers deforestation undertaken by any type of agent, including individuals, households and companies. Although it fails to specify the agent of deforestation, its strength is that it identifies the broad economic circumstances that either hinder or promote agriculturally driven deforestation.

The economy is divided into two main sectors, a modern sector (Y_M) comprised of the industrial (I) and services (S) sectors, and a traditional sector (Y_T) comprised of agricultural (A) and forestry (F) activities:

$$Aggregate\ output\ (Y) = \overbrace{A + F}^{Traditional\ sector,\ Y_T} + \underbrace{I + S}_{Modern\ sector,\ Y_M} \quad (5.1)$$

A country's aggregate production (Y) depends on three factors of production: land (T), labour (L) and capital (K). The traditional sector Y_T is characterized by using intensively labour and land in addition to capital, while the modern sector Y_M rather relies on capital and labour only (Lewis, 1954; Kirkpatrick and Barrientos, 2004). The two sector-specific production functions are:

$$Y_T = A_T\, f(K_T, L_T, T_A, T_F) \quad (5.2)$$

$$Y_M = A_M\, f(K_M, L_M) \quad (5.3)$$

Y_T is the aggregate output of the traditional sector, which depends on four factors of production: capital in agriculture (K_T), labour in the agriculture or forest sector (L_T), and land either in agriculture (T_A) or forest (T_F). The output of the modern sector (Y_M) depends on capital (K_M) and labour (L_M) in the sector. Technology (A) is assumed to be exogenous and specific to each sector. Both land and labour are assumed to be perfectly mobile between the agricultural and the forestry activities, meaning there are no labour or land market rigidities between these two sectors. Capital is assumed to be sector-specific, and the traditional and modern sectors of the economy are connected by exchanging one factor of production: labour. As such, our model corresponds to the traditional Ricardo-Viner trade model, except that we have an additional factor (land) in one of the sectors. A final, assumption is that labour is imperfectly mobile between the traditional and modern sector.

A key characteristic of dual economy models is that the marginal returns of each factor of production differs between sectors. If this condition is met, forest cover loss can be explained by two mechanisms, illustrated in Figure 1. If the marginal return of forestland is lower than the marginal return to agricultural land, deforestation is the result of a "stock adjustment" process: land is being allocated to its most productive use (agriculture). Second, if there is surplus labour in the traditional sector and imperfect labour mobility, an increase in labour supply in the traditional sector by, for example, rural population growth, will increase the demand for land, resulting in agricultural expansion and deforestation.

Similarly, forest cover gain occurs by two mechanisms. First, forest scarcity can drive increases in the marginal return of forestland, thus deforestation rates decrease (with an eventual forest recovery if the marginal return of forest becomes higher than the marginal return of agriculture). Second, if the marginal return to labour in the modern sector is higher than in the traditional sector and labour can move from the traditional sector to the modern one, then deforestation and agricultural expansion decreases as labour is pulled out of the traditional sector and into the modern sector. This mechanism illustrates the process of structural transformation during economic development, where the migration to urban areas and the "de–agrarization" of the national economy results in land abandonment and the natural regeneration of forests. These two mechanisms of forest recovery correspond to the "forest scarcity" path and the "modernization" path identified in the forest transition literature (Angelsen and Rudel, 2013).

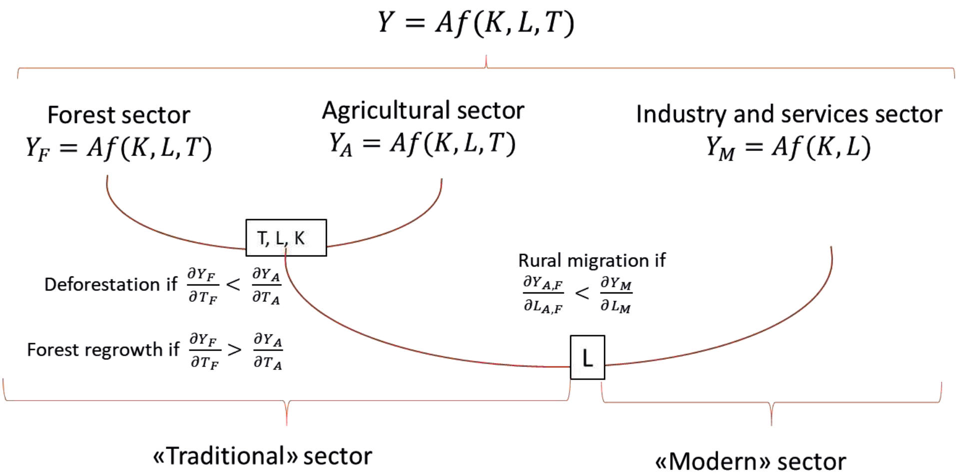


Figure 5.1 The process of forest cover gain in a three-sector economy. K is Capital, L is Labour, and T is Land.

In summary, national deforestation rates depend on how the marginal returns to each factor of production in the traditional and the modern sector, the marginal returns to agricultural land and forest land (Barbier et al., 2010), and the intersectoral factor mobility. I define the relative marginal return (RR) of each factor of production as follows:

$$RRT = \frac{MPT_A}{MPT_F} \tag{5.4}$$

$$RRL = \frac{MPL_M}{MPL_T} \tag{5.5}$$

$$RRK = \frac{MPK_M}{MPK_T} \tag{5.6}$$

MPT_A is the marginal product of agricultural land , MPT_F the marginal product of forest land, MPL_M the marginal product of labour in the modern sector, MPL_T the marginal product of labour in the traditional sector, while MPK_M and MPL_T are the marginal products of capital in the modern or traditional sector respectively[19]. Thus, *the immediate*

[19] Meaning $MPT_A = \frac{\partial Y_T}{\partial T_A}$; $MPT_F = \frac{\partial Y_T}{\partial T_F}$; $MPL_M = \frac{\partial Y_M}{\partial L_M}$; $MPL_T = \frac{\partial Y_T}{\partial L_T}$; $MPK_M = \frac{\partial Y_M}{\partial K_M}$; $MPK_T = \frac{\partial Y_T}{\partial K_M}$

drivers of deforestation can be defined as these ratios between the marginal returns of each factor of production. Accordingly, and for the purpose of the empirical analysis, I put forward three testable hypotheses:

H1: If RRT increases deforestation rates increase, as it implies that the marginal return of agricultural land is increasing relative to the marginal return of forestland.

H2: If RRL increases, deforestation rates decrease, given that the marginal return of labour in the modern sector relative to the marginal value product of labour in the traditional sector is greater (i.e., the "pull" factor out of the primary economy).

H3: RRK has no significant effect on deforestation, given that capital is sector specific.

The *underlying drivers* of deforestation are, in turn, the factors that determine the relative marginal returns (RR) of each factor of production. I focus on evaluating the effect of trade on the immediate drivers of deforestation. Trade and international trade policies is an often-studied subject in macroeconomic analysis of deforestation. Variables included to evaluate the effect of trade on deforestation can be the terms of trade and the exchange rates (Bulte and Barbier, 2005; Arcand et al., 2008; Faria and Almeida, 2016), openness to trade, measured as the shares of imports and exports as percent of GDP (Wolfersberger et al., 2015; Leblois et al., 2017). Considering only agricultural exports, trade is found to have significant effects on deforestation, but it is contingent on how much forest cover is left in the country (Leblois et al., 2017). External debt is also a variable included in some of studies, but have resulted in contradictory results (Culas, 2012; Köthke et al., 2013).

5.3 Empirical strategy

The empirical strategy consists of three steps. First, an aggregate production function was specified for each sector (modern and traditional), with the output elasticities of each factor of production are calculated and the marginal returns of each factor of production at the national level. The second step consists of evaluating how changes in the marginal returns of each factor of production in each sector relate to national deforestation rates. The third and final step consists of examining the underlying drivers of deforestation by examining how macroeconomic characteristics affect the immediate drivers of deforestation.

5.3.1 Estimation of the production functions

Two sector-specific production functions were specified as follows:

$$\frac{Y_T}{K_T} = A_T\, f(\frac{L_T}{K_T}, \frac{T_A}{K_T}, \frac{F}{K_T}) \tag{5.7}$$

$$\frac{Y_M}{K_M} = A_M\, f(\frac{L_M}{K_M}) \tag{5.8}$$

All inputs were divided by the capital input, thus assuming linear homogeneity (constant returns to scale) in production. Monotonicity in all inputs, which ensures positive elasticities, was imposed following Henningsen and Henning (2009).

Labour, land and aggregate output data were obtained from the World Development Indicators (WDI) database, from the year 1990 to 2015. Following ISIC classifications, the aggregate output of the traditional sector corresponds to the value added (constant 2010 USD) of forestry, hunting, and fishing activities, as well as cultivation of crops and livestock production. The aggregate output of the modern sector comprises value added (constant 2010 USD) in mining, manufacturing, construction, electricity, water, and gas. It also includes the value added in wholesale and retail trade (including hotels and restaurants), transport, and government, financial, professional, and personal services such as education, health care, and real estate services.

The data on agricultural land (T) refers to the land area that is arable, under permanent crops, and under permanent pastures. The forest area (F) is the land under natural or planted stands of trees, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens. The data on labour for the traditional sector is computed as the number of persons of working age who were engaged in any activity to produce goods in the traditional sector. Similarly, labour in the modern sector is the number of persons of working age who were engaged in any activity to produce industrial, mining, construction goods or provide services. The net capital stock data for the traditional sector was drawn from FAOSTAT, while capital stock for the modern sector was obtained from the Penn World Table Database (Feenstra et al., 2015). Both were converted to USD 2010 prices. Summary statistics of these variables and a complete description of data, the countries considered and the sources are available in Appendix D.

For each sector, I estimated a production function with constant and neutral technological change. Two particularities of the production functions estimated are first, that the production functions of the traditional sector include forest land as an input, and second, the use of a translog specification with constant returns to scale, rather than the standard Cobb-Douglas specification, as follows:

$$\ln y = \alpha_o + \sum_i \alpha_i \ln x_{c,i,t} + \frac{1}{2}\sum_i \sum_j \alpha_{ij} \ln x_{c,i,t} \ln x_{c,j,t} + \alpha_t t + u_{i,t} \quad (5.9)$$

A trans-log production function allows including interactions between the factors of production i and factor of production j with the coefficient α_{ij}, for each country and each year t of observation. The intercept α_o represents total factor productivity (TFP), α_i are the coefficients of each factor of production i. The linear time trend $\alpha_t t$ is added to account for technological change over time, i.e., the coefficient α_t is the rate of technological change per year. The output elasticities $e_{c,i,t}$ for each country c and factor of production i are calculated using the coefficients of equation (5.9):

$$e_{c,i,t} = \alpha_i + \sum_j \alpha_{ij} \ln x_{c,j,t} \tag{5.10}$$

The final sample includes 112 developed and developing countries (cf. Appendix D, Table 5.8). Countries that have a forest cover less than 10% of their land surface were not considered for the analysis, as well as small islands states and countries with insufficient data.

5.3.2 Estimation of the immediate drivers of deforestation

Assuming that factors of production are paid their marginal product (i.e., competitive market economy), it is possible to calculate the marginal returns (MP) of each factor of production of country c in year t. The country elasticities obtained from the production functions $e_{c,i,t}$ were multiplied by the average product (AP) of each factor of production, such that:

$$MP_{c,i,t} = \frac{\partial y_{c,t}}{\partial x_{c,i,t}} = \frac{y_{c,t}}{x_{c,i,t}} \frac{\partial \ln y_{c,t}}{\partial \ln x_{c,i,t.}} = AP_{c,i,t} \, (\alpha_i + \sum_j \alpha_{ij} \ln x_{c,j,t}) \tag{5.11}$$

Following, I proceeded to evaluate how the relative marginal returns of each factor of production (i.e., RRT and RRL) affect deforestation. For this purpose, I conducted panel data analysis with country and year fixed effects estimation. The econometric estimation was:

$$y_{c,t} = \alpha + \beta_1 (R\dot{R}T)_{c,t} + \beta_2 (R\dot{R}L)_{c,t} + \beta_3 (R\dot{R}K)_{c,t} + \beta_4 Z_{c,t} + \gamma_t + \delta_c + v_{c,t} \tag{5.12}$$

$y_{c,t}$ is our outcome variable (deforestation). To evaluate the sensitivity of our results to the definition of deforestation, I analyse three versions of the outcome variable. First, the deforestation *rate*, defined as the percent of total land area that is deforested each year. Second, the deforestation *level*, which is the total deforested area (in ha). The variables $R\dot{R}T$, $R\dot{R}L$, and $R\dot{R}K$ are the main independent variables, representing the growth of the relative returns of land, labour and capital across sectors. Variables γ_t represents time fixed effects and δ_c represents country fixed effects.

The vector $Z_{c,t}$ represents control variables: growth rate of GDP per capita and growth rate of population. To control for institutional characteristics and governance quality that might also affect deforestation rates (Barbier and Tesfaw, 2015), three institutional variables were included as control, from the World Governance Indicators: rule of law, political stability and regulatory quality. These measures are in units of a standard normal distribution, and range approximately from -2.5 to 2.5, where higher values correspond to better governance. Descriptions of these variables are found in Appendix D.

5.3.3 Estimation of the underlying drivers of deforestation

The main underlying drivers considered are indicators of international trade. The estimated equation for the underlying drivers of deforestation is as follows:

$$RR_{c,t} = \alpha + \beta_1 X_{c,t} + \beta_4 Z_{c,t} + \gamma_t + \delta_c + v_{c,t} \qquad (5.13)$$

Where $RR_{c,t}$ are the relative returns of each factor of production, $X_{c,t}$ are indicators of international trade, and $Z_{c,t}$ are the control variables. The analysis of the underlying drivers of deforestation focused on four indicators macroeconomic indicators: openness to trade (exports plus imports of goods and services, as % of GDP), agricultural imports (as % of merchandise imports), agricultural exports (as % of merchandise exports) and exchange rates. The outcome variables are RRT, RRL, and RRK.

5.3.4 Theory and data limitations

There are limitations to the theory and limitations in the data used to test the theory. Regarding the theory, to estimate the elasticities of each factor of production a general production function was estimated, which is a restrictive assumption if one considers that there is considerable cross-country variation in technologies. An extension would be to consider regional estimates of the production function. Another assumption is that one must assume that the factors of production are operating in a competitive economy. Regarding the data, an important limitation of the Hansen et al. (2013) data set is that it considers forest cover losses which includes clearing of tree crops such as oil palm. Thus, the data does not provide an exact indication of agricultural expansion. Furthermore, the data set only provides information on forest cover loss and not forest cover gain, which prevents evaluating net deforestation.

5.4 Results

5.4.1 Production functions and output elasticities

The results of the production function analysis indicate that all factors of production are significant inputs in the translog production function specification (Appendix D). The resulting average output elasticities of the traditional and modern sector are indicated in Table 5.1. The resulting elasticities show two important main messages. First, since the average output elasticities of each factor of production vary across sectors in each country, it corroborates the existence of dual economies, as specified in our theoretical framework. In general, agricultural land has higher returns than forest land, and labour and capital have higher returns in the modern sector than in the traditional sector.

Second, the output elasticities indicate that the characteristics of the dual economy vary across country income groups. Forest has a higher output elasticity than agricultural land in high income countries than low-income countries. In the modern sector, output elasticity of labour is higher for high income countries than low-income group. A detailed table of the average elasticities for each country is presented in Appendix D.

Table 5.1. Average output elasticities from the translog production function of each input in the traditional and modern sector, by income group. Standard deviations in parenthesis

	Output elasticity	**(1) Low Income**	**(2) Lower middle income**	**(3) Upper middle income**	**(4) High income**
Traditional sector	Agricultural land	0.57	0.48	0.44	0.28
		(0.07)	(0.07)	(0.02)	(0.09)
	Forest land	0.22	0.25	0.27	0.35
		(0.04)	(0.03)	(0.02)	(0.01)
	Agricultural labor	0.12	0.09	0.08	0.05
		(0.01)	(0.01)	(0.01)	(0.02)
	Agricultural capital	0.08	0.16	0.19	0.30
		(0.01)	(0.05)	(0.08)	(0.01)
Modern sector	Non-primary labor	0.35	0.42	0.47	0.55
		(0.05)	(0.06)	(0.05)	(0.04)
	Non primary capital	0.64	0.57	0.53	0.44
		(0.01)	(0.04)	(0.04)	(0.04)
	N	289	426	544	510

The relationship between output elasticities across sectors and income groups is further indicated in Figures 5.2 to 5.4. As income (GDP per capita) increases, the difference between output elasticities of forest land and agricultural land decrease. That is, as income increases, there is decreasing productivity of agricultural lands as compared to forest lands. This suggests that for low income countries, the incentives to deforest are highest. Labour shows the opposite trend: as the country income increases, there are increasing differences between the modern sector and the traditional sector (Figure 4).

The higher the income level, the bigger the difference between the output elasticity of labour in the modern sector as compared to the traditional sector. This is an indication of the low development of the modern sector in low income countries, which keeps marginal labour productivity low. Finally, regarding capital, the higher the country income level, the lower the difference between the output elasticity of capital in the modern sector as compared to the traditional sector (Figure 5). As with labour, this is an indication of the lower level of development of the industry and services sectors in low income countries, as well of a higher capital mobility in high income countries than in low income countries.

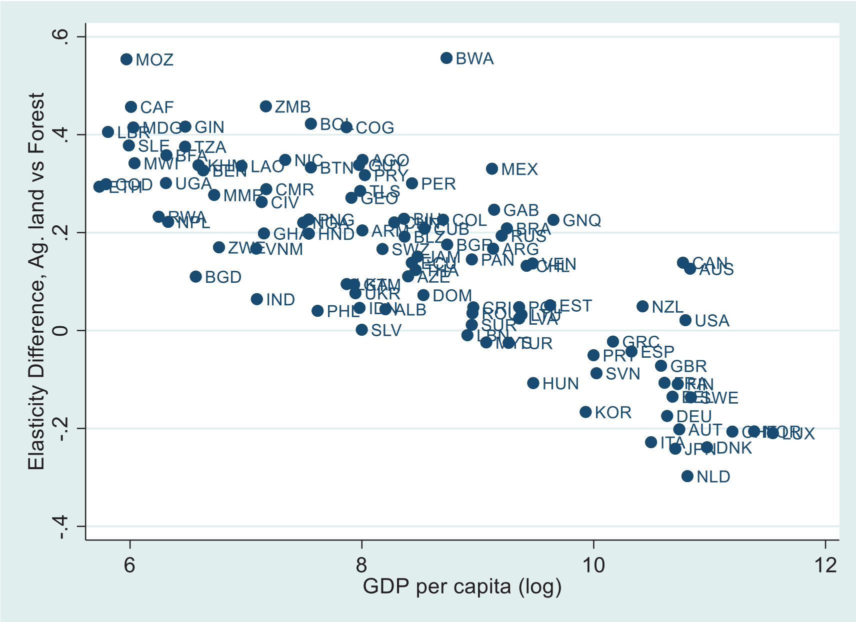


Figure 5.2. Difference in average output elasticity of agricultural land and forestland, per country.

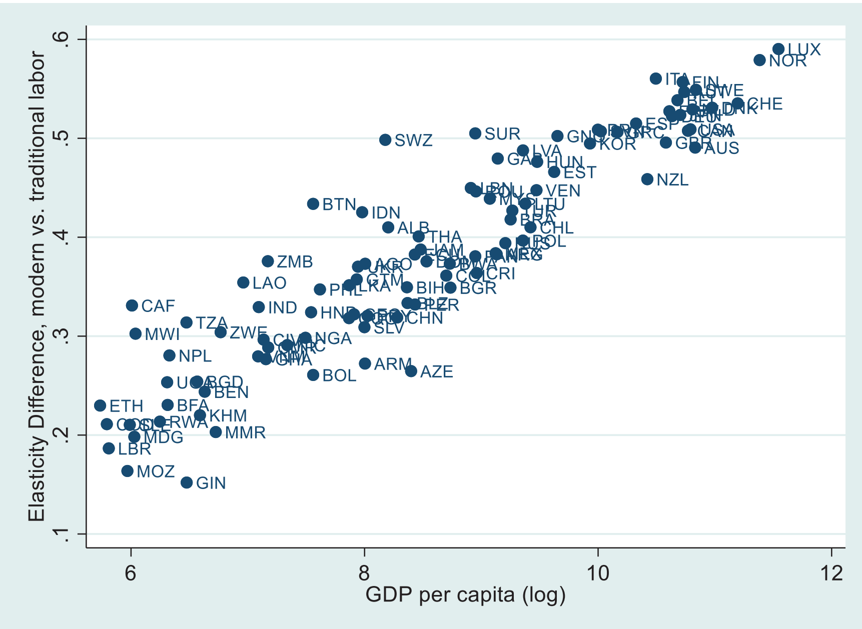


Figure 5.3. Difference in average output elasticity of labour in the modern sector and labour in the traditional sector, per country.

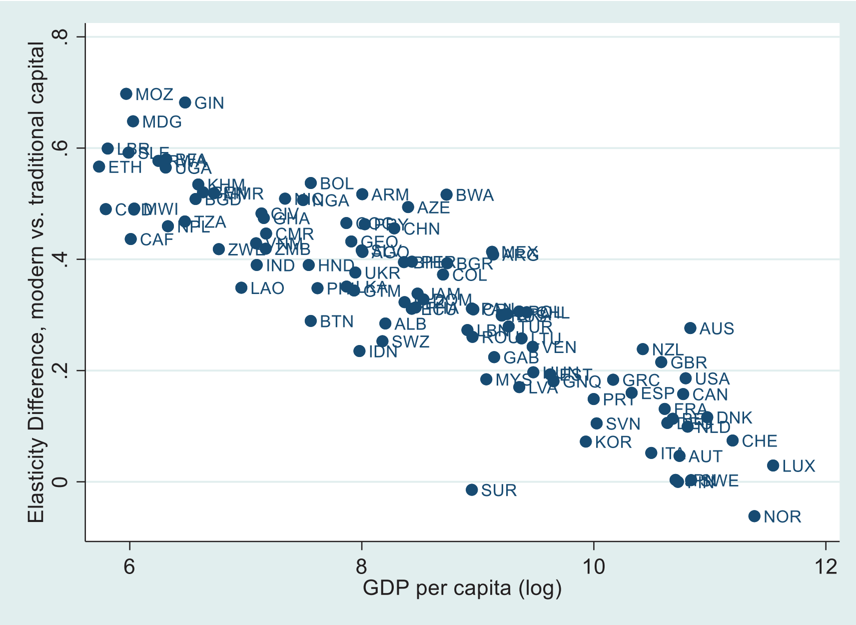


Figure 5.4. Difference in average output elasticity of capital in the modern sector and capital in the traditional sector, per country.

In summary, the results thus far provide evidence of the existence of a dual economy and indicate that the degree and nature of the duality depends on the income level of the country. In general, countries can increase aggregate output by the reallocation of factors of production, and compared to high income countries, low-income countries would benefit most from the reallocation of more forest land to agriculture, and more capital from the traditional sector to the modern sector.

5.4.2 Immediate drivers of deforestation

Results from the regression analysis of the immediate drivers of deforestation indicate that when controlling for economic growth and population density, only the growth rates in the relative returns of land across sectors (RRT) is a significant determinant of deforestation (Table 5.2). These results are consistent considering both the deforestation levels and the deforestation rate. There are no significant relationships to the deforestation growth rates. The results are robust to the inclusion of the institutional variables. Only the “rule of law” index is found to be significantly and positively related to deforestation, which is an indication that more secure contract enforcement and property rights creates safer investments and thus incentivize more agricultural expansion.

Table 5.2. Immediate drivers of deforestation

	Defor. rate (log)	*Defor. level (log)*	*Defor.rate (log)*	*Defor. level (log)*
	(1)	(2)	(3)	(4)
RRT growth	**0.785*****	**0.777*****	**0.879*****	**0.872*****
	(0.274)	**(0.274)**	**(0.291)**	**(0.291)**
RRL growth	0.142	0.142	0.160	0.160
	(0.116)	(0.116)	(0.122)	(0.122)
RRK growth	0.015	0.015	0.014	0.014
	(0.092)	(0.092)	(0.093)	(0.093)
GDPpc growth	-0.268	-0.268	-0.141	-0.140
	(0.354)	(0.353)	(0.371)	(0.371)
Population density growth	0.85***	0.85***	0.94***	0.94***
	(2.814)	(2.814)	(2.986)	(2.986)
Political stability			-0.000	-0.000
			(0.060)	(0.060)
Rule of law			0.279**	0.277**
			(0.134)	(0.134)
Regulatory quality			-0.065	-0.062
			(0.108)	(0.108)
Observations	1,211	1,211	1,121	1,121
R^2	0.011	0.011	0.018	0.018
Country FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
F Statistic	2.365**	2.330**	2.275**	2.242**

Note: ***p<0.01, **p<0.05, **p<0.01. Country clustered standard errors in parenthesis.

To further explore the variation in the observations and the effect of the immediate drivers of deforestation across income groups, we separate the analysis by income groups (Table 5.3). I find that the significant relationship between structural change indicator variables (i.e., the relative returns to each factor of production) only occur for low income and lower middle-income countries. In addition, an important observation is that growth in GDP per capita is not a significant determinant of deforestation. This suggests that the process of growth, rather than income level in general, is a more important determinant of deforestation.

Table 5.3. Fixed effects estimations of immediate drivers of deforestation, by income group.

	High income and upper middle income countries		Low income and lower middle income countries	
	Defor. rate (log)	*Defor. level (log)*	*Defor. rate (log)*	*Defor. level (log)*
	(1)	*(2)*	*(3)*	*(4)*
RRT growth	1.015	1.016	**1.087***	**1.059***
	(0.756)	(0.756)	**(0.584)**	**(0.584)**
RRL growth	0.061	0.060	-0.011	-0.011
	(0.193)	(0.193)	(0.232)	(0.232)
RRK growth	0.290	0.290	-0.075	-0.075
	(0.218)	(0.218)	(0.111)	(0.111)
GDPpc growth	-0.860*	-0.859*	0.721	0.726
	(0.517)	(0.517)	(0.578)	(0.578)
Population density growth	0.51	0.55	0.11*	0.10*
	(3.832)	(3.833)	(6.134)	(6.135)
Political stability	-0.064	-0.064	0.077	0.077
	(0.084)	(0.084)	(0.091)	(0.091)
Rule of law	0.018	0.017	0.572***	0.567***
	(0.188)	(0.188)	(0.195)	(0.195)
Regulatory quality	0.083	0.082	-0.325*	-0.319*
	(0.139)	(0.139)	(0.172)	(0.172)
Observations	696	696	425	425
Country FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
R^2	0.022	0.022	0.043	0.042
F Statistic	1.769*	1.767*	2.079**	2.032**

Note: ***p<0.01, **p<0.05, **p<0.01. Country clustered standard errors in parenthesis.

5.4.3 Underlying drivers of deforestation

The next step in the empirical strategy is to analyse the relationship between direct and indirect drivers of deforestation. Only openness to trade is significantly related to RRT (Table 5.4). Thus, opening to trade can have positive effects on structural transformation: it decreases the ratio between the marginal return of forest as compared to agricultural land (RRT), and in turn decreases deforestation. In other words, trade can indirectly reduce deforestation by decreasing the marginal returns of agriculture as compared to the marginal returns of forest land. Since RRT is the only variable that is significantly related to trade indicators, I conduct the regression with one subsample for high income and middle-income countries, and one for low income and lower middle-income countries (Table 4, column 7 and 8). Results indicate that it is only high income and upper middle-income countries that have a positive relationship between trade openness and changes in the relative returns of forest land as compared to agricultural land. This suggests that when opening to trade, these countries face lower agricultural returns or higher returns to forest land.

Table 5.4. Time and individual fixed effects estimations of underlying drivers of deforestation

	RRT growth	*RRL growth*	*RRK growth*	*RRT growth*	*RRL growth*	*RRK growth*	*RRT growth*	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7) Low income*	*(8) High income*
Openness to trade (% GDP)	-0.023^{**} (0.009)	0.001 (0.001)	0.00 (0.000)	-0.026^{**} (0.010)	0.001 (0.001)	0.001 (0.000)	0.006 (0.007)	-0.030^{**} (0.012)
Agricultural imports (% merchandise)	-0.012 (0.057)	0.002 (0.003)	0.003 (0.002)	-0.024 (0.060)	0.002 (0.003)	0.003 (0.002)	-0.005 (0.016)	0.146 (0.550)
Agricultural exports (% merchandise)	-0.016 (0.035)	-0.002 (0.002)	-0.002 (0.002)	-0.016 (0.039)	-0.001 (0.003)	-0.001 (0.002)	-0.007 (0.011)	-0.132 (0.148)
GDPpc growth	-0.242^{***} (0.089)	-0.268 (0.165)	-0.032 (0.133)	-0.248^{***} (0.093)	-0.341^{*} (0.176)	-0.131 (0.141)	0.047 (0.044)	0.118 (0.307)
Population density growth	-0.222^{***} (0.045)	0.074 (0.206)	0.673^{***} (0.167)	-0.219^{***} (0.047)	0.148 (0.218)	0.802^{***} (0.174)	-0.494^{***} (0.129)	-0.246^{***} (0.053)
Exchange rate	-0.000 (0.013)	0.001 (0.001)	0.001 (0.001)	-0.003 (0.015)	0.001 (0.001)	0.001 (0.001)	0.012 (0.009)	-0.008 (0.019)
Political stability				0.733 (0.546)	-0.025 (0.031)	-0.025 (0.025)	-0.071 (0.245)	1.282 (0.816)
Rule of law				-1.371 (1.259)	0.123^{*} (0.066)	0.067 (0.052)	0.227 (0.672)	-2.529 (1.697)
Regulatory quality				0.096 (1.018)	-0.055 (0.051)	-0.011 (0.040)	-1.311^{**} (0.649)	0.689 (1.256)
Observations	765	662	662	715	613	613	152	563
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.053	0.012	0.040	0.058	0.019	0.055	0.211	0.064
F Statistic	6.442^{***}	1.153	4.066^{***}	4.335^{***}	1.170	3.461^{***}	3.454^{***}	3.810^{***}

Note: $^{***}p<0.01$, $^{**}p<0.05$, $^{**}p<0.01$. Country clustered standard errors in parenthesis

5.5 Discussion

5.5.1 Dual economies and cross-country productivity differences

The results presented in this paper indicate that there are sectoral productivity differences across sectors, and the nature and degree of these sectoral differences varies according to country income levels. In particular, the marginal productivity of agricultural land is higher than forest land, and the marginal productivity of labour and capital is higher in the modern sector than in the traditional sector, which confirms the initial hypotheses. In general, labour productivity in the traditional sector is argued to be lower than in the modern sector (Gollin et al., 2014), which is consistent with our results.

Cross sectoral elasticity differences in capital and land diminish as national income increases, which is consistent with a process of structural adjustment along economic development. In the case of land, as income increases, land is reallocated to the most productive land use (agriculture), and thus the differences in the marginal productivity between forest land and agricultural land decrease. Regarding capital, the difference in the marginal return to capital in the modern as compared to the traditional sector is highest at early stages of economic development but decreases as income levels rise.

While cross sectoral elasticity differences in capital and land diminish as national income increases, I obtained an opposing trend with labour productivity: differences in cross sectoral labour productivity increase as GDP per capita increases. What this implies is that the “pull-factor” that induces labour to move from the traditional sector to the modern sector is greater in higher income countries than in low-income countries. This could be explained by a lower development level of the modern sector in low-income or lower middle-income countries, or stronger labour market rigidities to shift workers from the agricultural sector to the non-agricultural sector. The modern sector may require labour with more skills and education than the traditional sector, and this skill requirement difference is probably bigger in poor countries than in rich countries.

In addition, there are higher cross-country differences in the labour output elasticity of the modern sector than in the traditional sector. This contradicts other studies that have found that differences in cross country labour productivity are larger in agriculture than in the non-agricultural sector (Caselli, 2005; Restuccia et al., 2008; Lagakos and Waugh, 2013). To complement this analysis, alternative specifications of the production function and country specific analysis could help improve the estimations of factor productivity and the estimations of the direct drivers of deforestation.

5.5.2 Identifying the drivers of deforestation

The hypothesis that differences in land use values are the main drivers of deforestation is common in the literature (Barbier et al., 2010). Results indicate that the only relevant indicator of structural change that is significantly related to deforestation rates are changes in the marginal returns to agricultural land as compared to forest land, which confirms the initial hypothesis. On the other hand, contrary to what was hypothesized, there is no effect of changes in cross-sectoral marginal returns to labour. This could be explained by omitted variables that could control for labour market rigidities in the economy, and for human capital, such as education. Another source of bias could stem from data quality: in many developing countries the informal economy comprises a high proportion of the total labour force and is not registered in national data bases. Regarding the changes in the marginal return of capital across sectors, the results show no effect on deforestation, which is in line with the initial hypothesis.

Regarding the analysis of trade as an underlying driver, I found that more trade can decrease deforestation rates by diminishing the differences in the marginal returns of agricultural land as compared to forest land. This finding may relate to differences in agricultural policy. High-income countries have historically tended to support their agricultural sector through tariffs and direct support, while developing countries have taxed agriculture, as shown by classical work (Krueger et al., 1988). Opening to trade in high-income countries has entailed more alignment with comparative advantage and lower marginal profitability in agriculture. These results contradict previous studies that have found negative effects of trade on forest cover (Leblois et al., 2017).

5.6 Conclusion

Drawing from dual economy models of economic growth, the paper developed a theoretical framework to analyse the drivers of deforestation and disentangle the immediate drivers form the indirect drivers. The direct drivers are associated with three indicators of structural change. The results are straightforward and consistent with economic predictions: it is the competing land use value between forest and agriculture that is the main immediate driver of deforestation. The other two indicators of structural change, which regard the factors of production labour and capital, were not significant immediate drivers of deforestation. There is also supporting evidence that openness to trade can, indirectly, reduce deforestation levels by decreasing the marginal returns to agriculture as compared to forest land, but only for high income countries.

Acknowledgements.

This research is part of CIFOR's Global Comparative Study on REDD+ (www.cifor.org/gcs). The funding partners that have supported this research include the Norwegian Agency for Development Cooperation (Norad, Grant No. QZA-16/0110 & No. QZA-21/0124), International Climate Initiative (IKI) of the German Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU, Grant No. 20_III_108), and the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA) with financial support from the CGIAR Fund Donors. I thank Arild Angelsen, Ragnar Øygard and Arne Henningsen for useful advice and comments.

Appendix D

Table 5.5. Summary statistics of input data for estimation of production functions

	Mean	**S. Dev**	**Min**	**Max**	**N**	**Unit**
Agricultural GDP	15.8	53.6	0.06	710	2585	Thousand Million USD, current 2010
Industrial GDP	112	369	0.21	4180	2584	Thousand Million USD, current 2010
Services GDP	287	966	1.25	12600	2361	Thousand Million USD, current 2010
Agricultural Labor	7.60	31	0.002	35.8	2479	Million Individuals
Industrial Labor	5.05	19	0.003	23.5	2479	Million Individuals
Services Labor	9.5	26	0.1	42.0	2479	Million Individuals
Capital	1.9	5.6	<0.01	67	2361	Million USD, current 2010
Agricultural Capital	1.7	32	0	91	2773	Million USD, current 2010
Agricultural Land	253.6	755.5	3	5278.33	2763	Thousand Sq. km
Forest Land	259.3	892.4	10	8151.36	2773	Thousand Sq. km

Table 5.6Descriptions of variables for the production functions estimations

Variable	**Definition**
Agricultural land (sq. km)	Agricultural land refers to the land area that is arable, under permanent crops, and under permanent pastures. Arable land includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow. Land abandoned as a result of shifting cultivation is excluded. Land under permanent crops is land cultivated with crops that occupy the land for long periods and need not be replanted after each harvest, such as cocoa, coffee, and rubber. This category includes land under flowering shrubs, fruit trees, nut trees, and vines, but excludes land under trees grown for wood or timber. Permanent pasture is land used for five or more years for forage, including natural and cultivated crops. Source: World Development Indicators, retrieved January 2019
Agriculture, forestry, and fishing, value added (constant 2010 US$)	Agriculture corresponds to ISIC divisions 1-5 and includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3 or 4. Source: World Development Indicators, retrieved January 2019
Employment in agriculture (% of total employment) (modeled ILO estimate)	Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job, or to working-time arrangement. The agriculture sector consists of activities in agriculture, hunting, forestry and fishing, in accordance with division 1 (ISIC 2) or categories A-B (ISIC 3) or category A (ISIC 4). Source: World Development Indicators, retrieved January 2019
Employment in industry (% of total employment) (modeled ILO estimate)	Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job, or to working-time arrangement. The industry sector consists of mining and quarrying, manufacturing, construction, and public utilities (electricity, gas, and water), in accordance with divisions 2-5 (ISIC 2) or categories C-F (ISIC 3) or categories B-F (ISIC 4) Source: World Development Indicators, retrieved January 2019
Employment in services (% of total employment) (modeled ILO	Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job, or to working-time arrangement. The services sector consists of wholesale and retail trade and restaurants and hotels; transport, storage, and

estimate)	communications; financing, insurance, real estate, and business services; and community, social, and personal services, in accordance with divisions 6-9 (ISIC 2) or categories G-Q (ISIC 3) or categories G-U (ISIC 4). Source: World Development Indicators, retrieved January 2019
Forest area (sq. km)	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens. Source: World Development Indicators, retrieved January 2019
Industry (including construction), value added (constant 2010 US$)	Industry corresponds to ISIC divisions 10-45 and includes manufacturing (ISIC divisions 15-37). It comprises value added in mining, manufacturing (also reported as a separate subgroup), construction, electricity, water, and gas. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3. Data are in current U.S. dollars. Source: World Development Indicators, retrieved January 2019
Labor force, total	Labor force comprises people ages 15 and older who supply labor for the production of goods and services during a specified period. It includes people who are currently employed and people who are unemployed but seeking work as well as first-time job-seekers. Not everyone who works is included, however. Unpaid workers, family workers, and students are often omitted, and some countries do not count members of the armed forces. Labor force size tends to vary during the year as seasonal workers enter and leave. Source: World Development Indicators, retrieved January 2019
Services, value added (constant 2010 US$)	Services correspond to ISIC divisions 50-99. They include value added in wholesale and retail trade (including hotels and restaurants), transport, and government, financial, professional, and personal services such as education, health care, and real estate services. Also included are imputed bank service charges, import duties, and any statistical discrepancies noted by national compilers as well as discrepancies arising from rescaling. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The industrial origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3 or 4. Data are in constant 2010 U.S. dollars. Source: World Development Indicators, retrieved January 2019

Capital stock, modern sector	Capital stock at constant 2011 national prices (in mil. 2011US$) Source: Penn World Table, version 9.0
Capital stock, traditional sector	Capital formation is measured by the total value of a producer's acquisitions, less disposals, of fixed assets during the accounting period plus certain additions to the value of non- produced assets (such as subsoil assets or major improvements in the quantity, quality or productivity of land) realised by the productive activity of institutional units. It takes into account the consumption (depreciation) of fixed capital. It is calculated using the Perpetual Inventory Method from time series of Gross Fixed Capital Formation (GFCF). http://www.fao.org/faostat/en/#data/CS Source: FAOSTAT, consulted and downloaded August 2019
Rule of Law	Rule of Law captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e., ranging from approximately -2.5 to 2.5. Source: World Governance Indicators, retrieved April 2019
Regulatory Quality	Regulatory Quality captures perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e., ranging from approximately -2.5 to 2.5. Source: World Governance Indicators, retrieved April 2019
Political Stability	Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically-motivated violence, including terrorism. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e., ranging from approximately -2.5 to 2.5. Source: World Governance Indicators, retrieved April 2019

Table 5.7. Translog Production Functions with constant and neutral technological change

	GDP (log), Traditional sector (1)	GDP (log), Modern sector (2)
Ln Agricultural land (km^2)	0.594***	
	(0.044)	
Ln Forest land (km^2)	0.166***	
	(0.039)	
Ln Labor (total)	0.050**	0.425***
	(0.021)	(0.018)
(0.5 * *Ln* Agricultural land^2)	0.011	
	(0.026)	
(0.5 * *Ln* Forest land ^{2})	0.090***	
	(0.029)	
(0.5 * *Ln* Labor^2)	-0.006	-0.085***
	(0.009)	(0.019)
Ln Agricultural land * *Ln* Forest land	-0.053**	
	(0.025)	
Ln Labor * *Ln* Agricultural land	0.047***	
	(0.009)	
Ln Forest land * *Ln* Labor	-0.049***	
	(0.011)	
Year	0.359***	0.152***
	(0.014)	(0.011)
Observations	2,162	1,968
R^2	0.615	0.238
Adjusted R^2	0.592	0.195
F Statistic	325.575*** (df = 10; 2041)	194.092*** (df = 3; 1860)
Notes:	***p<0.01, **p<0.05, **p<0.01	

Table 5.8. Output table from the estimations of the production functions.

Country Code	Country Name	Traditional sector				Modern Sector	
		Agricultural land	Forest land	Labor	Capital	Labor	Capital
AGO	Angola	0,57	0,22	0,10	0,11	0,48	0,52
ALB	Albania	0,37	0,33	0,09	0,22	0,50	0,50
ARG	Argentina	0,48	0,31	0,08	0,13	0,46	0,54
ARM	Armenia	0,50	0,29	0,11	0,10	0,39	0,61
AUS	Australia	0,45	0,32	0,07	0,16	0,56	0,44
AUT	Austria	0,20	0,40	0,04	0,37	0,59	0,41
AZE	Azerbaijan	0,43	0,32	0,10	0,14	0,37	0,63
BEL	Belgium	0,26	0,39	0,06	0,29	0,59	0,41
BEN	Benin	0,55	0,22	0,11	0,13	0,35	0,65
BFA	Burkina Faso	0,58	0,23	0,12	0,07	0,35	0,65
BGD	Bangladesh	0,44	0,33	0,12	0,12	0,37	0,63
BGR	Bulgaria	0,46	0,28	0,09	0,17	0,44	0,56
BIH	Bosnia and Herzegovina	0,49	0,26	0,10	0,16	0,44	0,56
BLZ	Belize	0,42	0,23	0,07	0,28	0,40	0,60
BOL	Bolivia	0,61	0,19	0,11	0,10	0,37	0,63
BRA	Brazil	0,46	0,26	0,08	0,20	0,50	0,50
BRN	Brunei Darussalam	0,20	0,31	0,02	0,47	0,61	0,39
BTN	Bhutan	0,53	0,19	0,09	0,18	0,53	0,47
BWA	Botswana	0,72	0,17	0,13	-0,02	0,50	0,50
CAF	Central African Republic	0,61	0,16	0,11	0,12	0,44	0,56
CAN	Canada	0,40	0,26	0,05	0,29	0,56	0,44
CHE	Switzerland	0,20	0,41	0,05	0,34	0,58	0,42
CHL	Chile	0,42	0,29	0,08	0,21	0,49	0,51
CHN	China	0,50	0,28	0,11	0,12	0,43	0,57
CIV	Cote d'Ivoire	0,52	0,26	0,11	0,11	0,40	0,60
CMR	Cameroon	0,51	0,22	0,10	0,17	0,39	0,61
COD	Congo, Dem. Rep.	0,50	0,20	0,09	0,21	0,30	0,70
COG	Congo, Rep.	0,60	0,18	0,10	0,11	0,42	0,58
COL	Colombia	0,48	0,25	0,09	0,18	0,45	0,55
CRI	Costa Rica	0,36	0,31	0,07	0,26	0,43	0,57
CUB	Cuba	0,49	0,28	0,10	0,13		
CYP	Cyprus	0,24	0,37	0,05	0,33	0,54	0,46
DEU	Germany	0,23	0,40	0,05	0,33	0,57	0,43

DNK	Denmark	0,20	0,44	0,05	0,30	0,58	0,42
DOM	Dominican Republic	0,39	0,32	0,08	0,21	0,46	0,54
ECU	Ecuador	0,42	0,28	0,08	0,23	0,46	0,54
ERI	Eritrea	0,72	0,18	0,15	-0,05		
ESP	Spain	0,32	0,36	0,06	0,26	0,58	0,42
EST	Estonia	0,35	0,30	0,06	0,28	0,52	0,48
ETH	Ethiopia	0,54	0,25	0,12	0,08	0,35	0,65
FIN	Finland	0,22	0,33	0,03	0,42	0,58	0,42
FRA	France	0,28	0,38	0,06	0,29	0,58	0,42
GAB	Gabon	0,48	0,23	0,07	0,22	0,55	0,45
GBR	United Kingdom	0,32	0,39	0,07	0,22	0,57	0,43
GEO	Georgia	0,51	0,24	0,10	0,14	0,43	0,57
GHA	Ghana	0,47	0,28	0,10	0,15	0,38	0,62
GIN	Guinea	0,62	0,21	0,13	0,04	0,28	0,72
GNQ	Equatorial Guinea	0,45	0,23	0,08	0,24	0,59	0,41
GRC	Greece	0,33	0,36	0,07	0,24	0,58	0,42
GTM	Guatemala	0,40	0,30	0,09	0,21	0,44	0,56
GUY	Guyana	0,52	0,18	0,07	0,22		
HND	Honduras	0,46	0,26	0,09	0,20	0,41	0,59
HUN	Hungary	0,28	0,39	0,06	0,27	0,54	0,46
IDN	Indonesia	0,35	0,31	0,08	0,26	0,50	0,50
IND	India	0,39	0,33	0,10	0,18	0,43	0,57
ITA	Italy	0,19	0,42	0,04	0,34	0,60	0,40
JAM	Jamaica	0,44	0,29	0,10	0,18	0,48	0,52
JPN	Japan	0,14	0,38	0,03	0,44	0,55	0,45
KHM	Cambodia	0,55	0,21	0,11	0,14	0,33	0,67
KOR	Korea, Rep.	0,20	0,37	0,05	0,39	0,54	0,46
LAO	Lao PDR	0,52	0,18	0,09	0,20	0,45	0,55
LBN	Lebanon	0,36	0,37	0,08	0,20	0,53	0,47
LBR	Liberia	0,60	0,19	0,11	0,10	0,30	0,70
LKA	Sri Lanka	0,40	0,30	0,09	0,21	0,44	0,56
LTU	Lithuania	0,36	0,33	0,07	0,24	0,51	0,49
LUX	Luxembourg	0,20	0,41	0,04	0,34	0,63	0,37
LVA	Latvia	0,34	0,32	0,06	0,28	0,55	0,45
MDG	Madagascar	0,63	0,22	0,13	0,02	0,33	0,67
MEX	Mexico	0,56	0,23	0,11	0,10	0,49	0,51
MKD	Macedonia, FYR	0,45	0,28	0,09	0,18	0,45	0,55
MMR	Myanmar	0,50	0,22	0,10	0,18	0,30	0,70

MNE	Montenegro	0,42	0,27	0,07	0,23	0,47	0,53
MOZ	Mozambique	0,71	0,15	0,14	0,00	0,30	0,70
MWI	Malawi	0,57	0,23	0,12	0,08	0,43	0,57
MYS	Malaysia	0,30	0,32	0,06	0,32	0,50	0,50
NGA	Nigeria	0,51	0,29	0,12	0,08	0,42	0,58
NIC	Nicaragua	0,57	0,23	0,11	0,09	0,40	0,60
NLD	Netherlands	0,17	0,46	0,05	0,32	0,58	0,42
NOR	Norway	0,16	0,36	0,01	0,47	0,59	0,41
NPL	Nepal	0,48	0,26	0,11	0,15	0,39	0,61
NZL	New Zealand	0,37	0,32	0,06	0,24	0,52	0,48
PAN	Panama	0,42	0,27	0,08	0,23	0,46	0,54
PER	Peru	0,51	0,21	0,09	0,18	0,42	0,58
PHL	Philippines	0,37	0,33	0,09	0,22	0,44	0,56
PNG	Papua New Guinea	0,42	0,20	0,07	0,31		
POL	Poland	0,38	0,33	0,08	0,22	0,48	0,52
PRI	Puerto Rico	0,26	0,35	0,05	0,34		
PRT	Portugal	0,30	0,35	0,06	0,28	0,57	0,43
PRY	Paraguay	0,55	0,23	0,10	0,12	0,42	0,58
ROU	Romania	0,37	0,34	0,08	0,21	0,53	0,47
RUS	Russian Federation	0,44	0,25	0,07	0,24	0,46	0,54
RWA	Rwanda	0,51	0,28	0,13	0,08	0,34	0,66
SLE	Sierra Leone	0,59	0,21	0,12	0,08	0,33	0,67
SLV	El Salvador	0,36	0,36	0,09	0,18	0,40	0,60
SOM	Somalia	0,67	0,22	0,14	-0,02		
SRB	Serbia	0,39	0,31	0,08	0,21	0,51	0,49
SUR	Suriname	0,25	0,24	0,01	0,50	0,51	0,49
SVN	Slovenia	0,26	0,35	0,05	0,34	0,56	0,44
SWE	Sweden	0,21	0,34	0,02	0,43	0,57	0,43
SWZ	Eswatini	0,46	0,29	0,10	0,15	0,60	0,40
THA	Thailand	0,42	0,30	0,09	0,19	0,50	0,50
TLS	Timor-Leste	0,51	0,23	0,10	0,17		
TUR	Turkey	0,34	0,37	0,08	0,21	0,51	0,49
TZA	Tanzania	0,58	0,20	0,11	0,10	0,43	0,57
UGA	Uganda	0,56	0,26	0,13	0,05	0,38	0,62
UKR	Ukraine	0,41	0,33	0,09	0,16	0,46	0,54
USA	United States	0,36	0,34	0,06	0,25	0,57	0,43
VEN	Venezuela, RB	0,41	0,28	0,07	0,24	0,52	0,48
VNM	Vietnam	0,44	0,27	0,10	0,19	0,38	0,62

ZMB	Zambia	0,63	0,17	0,11	0,09	0,49	0,51
ZWE	Zimbabwe	0,45	0,28	0,09	0,19	0,40	0,60

References

ABADIE, A. & IMBENS, G. W. 2016. Matching on the Estimated Propensity Score. *Econometrica,* 84**,** 781-807. 10.3982/ecta11293.

ABBINK, K., GANGADHARAN, L., HANDFIELD, T. & THRASHER, J. 2017. Peer punishment promotes enforcement of bad social norms. *Nature communications,* 8**,** 1-8.

ADHIKARI, B. & LOVETT, J. C. 2006. Institutions and collective action: Does heterogeneity matter in community-based resource management? *The Journal of Development Studies,* 42**,** 426-445. 10.1080/00220380600576201.

AGGARWAL, A. & BROCKINGTON, D. 2020. Reducing or creating poverty? Analyzing livelihood impacts of forest carbon projects with evidence from India. *Land Use Policy,* 95. 10.1016/j.landusepol.2020.104608.

AGRAWAL, A. 2001. Common Property Institutions and Sustainable Governance of Resources. *World Development,* 29**,** 1649-1672.

AI, C. & NORTON, E. C. 2003. Interaction terms in logit and probit models. *Economics Letters,* 80**,** 123-129. 10.1016/s0165-1765(03)00032-6.

ALBRECHT, F., KUBE, S. & TRAXLER, C. 2018. Cooperation and norm enforcement - The individual-level perspective. *Journal of Public Economics,* 165**,** 1-16. 10.1016/j.jpubeco.2018.06.010.

ALIX-GARCIA, J. 2008. An exploration of the positive effect of inequality on common property forests. *Journal of Development Economics,* 87**,** 92-105. 10.1016/j.jdeveco.2007.12.009.

ALIX-GARCIA, J. & WOLFF, H. 2014. Payment for Ecosystem Services from Forests. *Annual Review of Resource Economics,* 6**,** 361-380. 10.1146/annurev-resource-100913-012524.

ALIX-GARCIA, J. M., MCINTOSH, C., SIMS, K. R. & WELCH, J. R. 2013. The Ecological Footprint of Poverty Alleviation: Evidence from Mexico's Oportunidades Program. *The Review of Economics and Statistics,* 95.

ALIX-GARCIA, J. M., SIMS, K. R., OROZCO-OLVERA, V. H., COSTICA, L. E., MEDINA, J. D. F. & MONROY, S. R. J. P. O. T. N. A. O. S. 2018. Payments for environmental services supported social capital while increasing land management. 115**,** 7016-7021.

ALIX-GARCIA, J. M., SIMS, K. R. E. & YAÑEZ-PAGANS, P. 2015. Only One Tree from Each Seed? Environmental Effectiveness and Poverty Alleviation in Mexico's Payments for Ecosystem Services Program. *American Economic Journal: Economic Policy,* 7**,** 1-40. 10.1257/pol.20130139.

ALMEIDA-LEÑERO, L., REVOLLO-FERNÁNDEZ, D., CARO-BORRERO, A., RUIZ-MALLÉN, I., CORBERA, E., MAZARI-HIRIART, M. & FIGUEROA, F. 2017. Not the same for everyone: Community views of Mexico's payment for environmental services programmes. *Environmental Conservation,* 44**,** 201-211. 10.1017/s0376892916000564.

ALMÅS, I., CAPPELEN, A. W., SORENSEN, E. O. & TUNGODDEN, B. 2010. Fairness and the development of inequality acceptance. *Science,* 328**,** 1176-8. 10.1126/science.1187300.

ALTHELIA 2018. Althelia Impact Report 2018. https://althelia.com/wp-content/uploads/2018/09/Althelia_impact-report_2018.pdf

ANDERIES, J. M., JANSSEN, M. A., BOUSQUET, F., CARDENAS, J.-C., CASTILLO, D., LOPEZ, M.-C., TOBIAS, R., VOLLAN, B. & WUTICH, A. 2011. The challenge of understanding decisions in experimental studies of common pool resource governance. *Ecological Economics,* 70**,** 1571-1579. 10.1016/j.ecolecon.2011.01.011.

ANDERSSON, K. & AGRAWAL, A. 2011. Inequalities, institutions, and forest commons. *Global Environmental Change,* 21**,** 866-875. 10.1016/j.gloenvcha.2011.03.004.

ANDERSSON, K., BENAVIDES, J. P. & LEÓN, R. 2014. Institutional diversity and local forest governance. *Environmental Science & Policy,* 36**,** 61-72. 10.1016/j.envsci.2013.07.009.

ANDERSSON, K. P., COOK, N. J., GRILLOS, T., LOPEZ, M. C., SALK, C. F., WRIGHT, G. D. & MWANGI, E. 2018a. Experimental evidence on payments for forest commons conservation. *Nature Sustainability,* 1**,** 128-135. 10.1038/s41893-018-0034-z.

ANDERSSON, K. P., SMITH, S. M., ALSTON, L. J., DUCHELLE, A. E., MWANGI, E., LARSON, A. M., DE SASSI, C., SILLS, E. O., SUNDERLIN, W. D. & WONG, G. Y. 2018b. Wealth and the distribution of benefits from tropical forests: Implications for REDD+. *Land Use Policy,* 72**,** 510-522. 10.1016/j.landusepol.2018.01.012.

ANGELSEN, A. 1999. Agricultural expansion and deforestation: modelling the impact of population, market forces and property rights. *Journal of Development Economics,* 58**,** 185-218.

ANGELSEN, A. 2017. REDD+ as Result-based Aid: General Lessons and Bilateral Agreements of Norway. *Review of Development Economics,* 21**,** 237-264. 10.1111/rode.12271.

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**,** 718-720. 10.1111/cobi.12933.

ANGELSEN, A., JAGGER, P., BABIGUMIRA, R., BELCHER, B., HOGARTH, N. J., BAUCH, S., BORNER, J., SMITH-HALL, C. & WUNDER, S. 2014. Environmental Income and Rural Livelihoods: A Global-Comparative Analysis. *World Dev,* 64**,** S12-S28. 10.1016/j.worlddev.2014.03.006.

ANGELSEN, A. & KAIMOWITZ, D. J. T. W. B. R. O. 1999. Rethinking the causes of deforestation: lessons from economic models. 14**,** 73-98.

ANGELSEN, A., LARSEN, H. O. & OLSEN, C. S. 2012. *Measuring livelihoods and environmental dependence: Methods for research and fieldwork*, Routledge.

ANGELSEN, A., MARTIUS, C., DE SY, V., DUCHELLE, A., LARSON, A. & PHAM, T. T. 2018. Introduction: REDD+ enters its second decade. *In:* ANGELSEN, A., MARTIUS, C., DE SY, V., DUCHELLE, A., LARSON, A. & PHAM, T. T. (eds.) *Transforming REDD+: Lessons and new directions.* Bogor, Indonesia: CIFOR.

ANGELSEN, A. & RUDEL, T. K. 2013. Designing and Implementing Effective REDD + Policies: A Forest Transition Approach. *Review of Environmental Economics and Policy,* 7**,** 91-113. 10.1093/reep/res022.

ANGELSEN, A. & WERTZ-KANOUNNIKOFF, S. 2008. Chapter 2. What are the key design issues for REDD and the criteria for assessing options? *In:* ANGELSEN, A. (ed.) *Moving ahead with REDD: issues, options and implications.* Bogor, Indonesia: CIFOR.

ARCAND, J.-L., GUILLAUMONT, P. & JEANNENEY, S. G. 2008. Deforestation and the real exchange rate. *Journal of Development Economics,* 86**,** 242-262. 10.1016/j.jdeveco.2007.02.004.

ARELLANO, M. & BOND, S. J. T. R. O. E. S. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. 58**,** 277-297.

ARNDT, C. & OMAN, C. 2006. Uses and Abuses of Governance Indicators. . Paris, France: Development Centre Studies: OECD Publishing,

ARRIAGADA, R., VILLASEÑOR, A., RUBIANO, E., COTACACHI, D. & MORRISON, J. 2018. Analysing the impacts of PES programmes beyond economic rationale: Perceptions of ecosystem services provision associated to the Mexican case. *Ecosystem Services,* 29**,** 116-127. 10.1016/j.ecoser.2017.12.007.

ASNER, G. P., KNAPP, D. E., MARTIN, R. E., TUPAYACHI, R., ANDERSON, C. B., MASCARO, J., SINCA, F., CHADWICK, K. D., HIGGINS, M., FARFAN, W., LLACTAYO, W. & SILMAN, M. R. 2014. Targeted carbon conservation at national scales with high-resolution monitoring. *Proc Natl Acad Sci U S A,* 111**,** E5016-22. 10.1073/pnas.1419550111.

ATHEY, S. & IMBENS, G. W. 2017. The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives,* 31**,** 3-32. 10.1257/jep.31.2.3.

ATMADJA, S., ARWIDA, S., MARTIUS, C. & PHAM, T. 2018. Financing REDD+: A transaction among equals, or an uneven playing field? *In:* ANGELSEN, A., MARTIUS, C., DE SY, V., DUCHELLE, A. E., LARSON, A. & PHAM, T. T. (eds.) *Transforming REDD+: Lessons and new directions.*
AUSTIN, K. G., GONZÁLEZ-ROGLICH, M., SCHAFFER-SMITH, D., SCHWANTES, A. M. & SWENSON, J. J. 2017. Trends in size of tropical deforestation events signal increasing dominance of industrial-scale drivers *Environmental Research Letters,* 12. 10.1088/1748-9326/aa7760.
AUTHELET, M., SUBERVIE, J., MEYFROIDT, P., ASQUITH, N. & EZZINE-DE-BLAS, D. 2021. Economic, pro-social and pro-environmental factors influencing participation in an incentive-based conservation program in Bolivia. *World Development,* 145. 10.1016/j.worlddev.2021.105487.
BALAFOUTAS, L., NIKIFORAKIS, N. & ROCKENBACH, B. 2016. Altruistic punishment does not increase with the severity of norm violations in the field. *Nat Commun,* 7**,** 13327. 10.1038/ncomms13327.
BALAND, J.-M. & PLATTEAU, J.-P. 1996. *Halting degradation of natural resources: is there a role for rural communities?*, Food and Agriculture Organization (FAO).
BALAND, J. M. & PLATTEAU, J. P. 1999. The Ambiguous Impact of Inequality on Local Resource Management. *World Development,* 27**,** 773-788.
BANERJEE, A. V. & DUFLO, E. 2005. Growth theory through the lens of development economics. *In:* AGHION, P. & DURLAUF, S. N. (eds.) *Handbook of economic growth.*
BARBIER, E. B. & BUGAS, J. S. 2014. Structural change, marginal land and economic development in Latin America and the Caribbean. *Lat Am Econ Rev,* 23.
BARBIER, E. B., BURGESS, J. C. & GRAINGER, A. 2010. The forest transition: Towards a more comprehensive theoretical framework. *Land Use Policy,* 27**,** 98-107. 10.1016/j.landusepol.2009.02.001.
BARBIER, E. B., DELACOTE, P. & WOLFERSBERGER, J. 2017. The economic analysis of the forest transition: A review. *Journal of Forest Economics,* 27**,** 10-17. 10.1016/j.jfe.2017.02.003.
BARBIER, E. B. & TESFAW, A. 2015. Explaining forest transitions: The role of governance. *Ecological Economics,* 119**,** 252-261. 10.1016/j.ecolecon.2015.09.010.
BEGAZO CURIE, K., MERTENS, K. & VRANKEN, L. 2021. Tenure regimes and remoteness: When does forest income reduce poverty and inequality? A case study from the Peruvian Amazon. *Forest Policy and Economics,* 128. 10.1016/j.forpol.2021.102478.
BENNETT, A., RAVIKUMAR, A. & PALTÁN, H. 2018. The Political Ecology of Oil Palm Company-Community partnerships in the Peruvian Amazon: Deforestation consequences of the privatization of rural development. *World Development,* 109**,** 29-41. 10.1016/j.worlddev.2018.04.001.
BENYISHAY, A., HEUSER, S., RUNFOLA, D. & TRICHLER, R. 2017. Indigenous land rights and deforestation: Evidence from the Brazilian Amazon. *Journal of Environmental Economics and Management,* 86**,** 29-47. 10.1016/j.jeem.2017.07.008.
BERNHARD, H., FEHR, E. & FISCHBACHER, U. 2006. Group Affiliation and Altruistic Norm Enforcement. *American Economic Review,* 96**,** 217-221.
BINSWANGER, H. 1981. Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India. *The Economics Journal,* 91.
BLACKMAN, A., CORRAL, L., LIMA, E. S. & ASNER, G. P. 2017. Titling indigenous communities protects forests in the Peruvian Amazon. *Proc Natl Acad Sci U S A,* 114**,** 4123-4128. 10.1073/pnas.1603290114.
BLACKMAN, A., PFAFF, A. & ROBALINO, J. 2015. Paper park performance: Mexico's natural protected areas in the 1990s. *Global Environmental Change,* 31**,** 50-61. 10.1016/j.gloenvcha.2014.12.004.

BLANCO, E., LOPEZ, M. C. & WALKER, J. M. 2016. The Opportunity Costs of Conservation with Deterministic and Probabilistic Degradation Externalities. *Environmental and Resource Economics,* 64**,** 255-273. 10.1007/s10640-014-9868-7.
BLUNDO-CANTO, G., BAX, V., QUINTERO, M., CRUZ-GARCIA, G. S., GROENEVELD, R. A. & PEREZ-MARULANDA, L. 2018. The Different Dimensions of Livelihood Impacts of Payments for Environmental Services (PES) Schemes: A Systematic Review. *Ecological Economics,* 149**,** 160-183. 10.1016/j.ecolecon.2018.03.011.
BONAN, G. B. 2008. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. 320**,** 1444-1449. doi:10.1126/science.1155121.
BOOSEY, L. & ISAAC, M. R. 2016. Asymmetric network monitoring and punishment in public goods experiments. *Journal of Economic Behavior & Organization,* 132**,** 26-41. 10.1016/j.jebo.2016.09.015.
BOS, A. B., DUCHELLE, A. E., ANGELSEN, A., AVITABILE, V., SY, V. D., HEROLD, M., JOSEPH, S., SASSI, C. D., SILLS, E. O., SUNDERLIN, W. D. & WUNDER, S. 2017. Comparing methods for assessing the effectiveness of subnational REDD+ initiatives. *Environmental Research Letters,* 12. 10.1088/1748-9326/aa7032.
BOWLES, S. & GINTIS, H. J. N. 2002. Homo reciprocans. 415**,** 125-127.
BOYD, R., GINTIS, H. & BOWLES, S. 2010. Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare. *Science,* 328.
BROUWER, R. O. Y., TESFAYE, A. & PAUW, P. 2011. Meta-analysis of institutional-economic factors explaining the environmental performance of payments for watershed services. *Environmental Conservation,* 38**,** 380-392. 10.1017/s0376892911000543.
BRUHIN, A., JANIZZI, K. & THÖNI, C. 2020. Uncovering the heterogeneity behind cross-cultural variation in antisocial punishment. *Journal of Economic Behavior & Organization,* 180**,** 291-308. 10.1016/j.jebo.2020.10.005.
BULTE, E. H. & BARBIER, E. B. 2005. Trade and Renewable Resources in a Second Best World: an overview. *Environmental and Resource Economics,* 30.
BUNTAINE, M. T. & DANIELS, B. 2020. Combining bottom-up monitoring and top-down accountability: A field experiment on managing corruption in Uganda. *Research & Politics,* 7. 10.1177/2053168020934350.
BUSCH, J. & ENGELMANN, J. 2017. Cost-effectiveness of reducing emissions from tropical deforestation, 2016–2050. *Environmental Research Letters,* 13. 10.1088/1748-9326/aa907c.
BUSCH, J. & FERRETTI-GALLON, K. 2017. What Drives Deforestation and What Stops It? A Meta-Analysis. *Review of Environmental Economics and Policy,* 11**,** 3-23. 10.1093/reep/rew013.
BUSCH, J., FERRETTI-GALLON, K., ENGELMANN, J., WRIGHT, M., AUSTIN, K. G., STOLLE, F., TURUBANOVA, S., POTAPOV, P. V., MARGONO, B., HANSEN, M. C. & BACCINI, A. 2015. Reductions in emissions from deforestation from Indonesia's moratorium on new oil palm, timber, and logging concessions. *Proc Natl Acad Sci U S A,* 112**,** 1328-33. 10.1073/pnas.1412514112.
BÖRNER, J., BAYLIS, K., CORBERA, E., EZZINE-DE-BLAS, D., HONEY-ROSÉS, J., PERSSON, U. M. & WUNDER, S. 2017. The Effectiveness of Payments for Environmental Services. *World Development,* 96**,** 359-374. 10.1016/j.worlddev.2017.03.020.
BÖRNER, J., SCHULZ, D., WUNDER, S. & PFAFF, A. 2020. The Effectiveness of Forest Conservation Policies and Programs. *Annual Review of Resource Economics,* 12**,** 45-64. 10.1146/annurev-resource-110119-025703.
BÖRNER, J., WUNDER, S., WERTZ-KANOUNNIKOFF, S., HYMAN, G. & NASCIMENTO, N. 2014. Forest law enforcement in the Brazilian Amazon: Costs and income effects. *Global Environmental Change,* 29**,** 294-305. 10.1016/j.gloenvcha.2014.04.021.

CAMERER, C. F. 2011. The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. *In:* FRECHETTE, G. R. & SCHOTTER, A. (eds.) *Handbook of Experimental Economic Methodology.* New York, USA: Oxford University Press.

CARAVAGGIO, N. 2020a. Economic growth and the forest development path: A theoretical re-assessment of the environmental Kuznets curve for deforestation. *Forest Policy and Economics,* 118. 10.1016/j.forpol.2020.102259.

CARAVAGGIO, N. 2020b. A global empirical re-assessment of the Environmental Kuznets curve for deforestation. *Forest Policy and Economics,* 119. 10.1016/j.forpol.2020.102282.

CARDENAS, C., STRANDLUND, J. & WILLIS, C. 2000. Local Environmental Control and Institutional Crowding out. *World Development,* 28.

CARDENAS, C., STRANDLUND, J. & WILLIS, C. 2002. Economic inequality and burden-sharing in the provision of local environmental quality. *Ecological Economics*, 379–395.

CARDENAS, J.-C. 2003. Real wealth and experimental cooperation: experiments in the field lab. *Journal of Development Economics,* 70, 263-289. 10.1016/s0304-3878(02)00098-6.

CARDENAS, J.-C. 2004. Norms from outside and from inside: an experimental analysis on the governance of local ecosystems. *Forest Policy and Economics,* 6, 229-241. 10.1016/j.forpol.2004.03.006.

CARDENAS, J.-C. & OSTROM, E. 2004. What do people bring into the game? Experiments in the field about cooperation in the commons. *Agricultural Systems,* 82, 307-326. 10.1016/j.agsy.2004.07.008.

CARDENAS, J. C. & CARPENTER, J. 2008. Behavioural Development Economics: Lessons from Field Labs in the Developing World. *The Journal of Development Studies,* 44, 311-338. 10.1080/00220380701848327.

CARPENTER, J., VERHOOGEN, E. & BURKS, S. J. E. L. 2005. The effect of stakes in distribution experiments. 86, 393-398.

CASELLI, F. 2005. Chapter 9 Accounting for Cross-Country Income Differences. *In:* AGHION, P. & DURLAUF, S. N. (eds.) *Handbook of Economic Growth.*

CASON, T. & GANGADHARAN, L. 2015. Promoting cooperation in nonlinear social dilemmas through peer punishment. *Experimental Economics,* 18, 66-88. 10.1007/s10683-014-9393-0.

CASON, T. N. & GANGADHARAN, L. 2013. Empowering neighbors versus imposing regulations: An experimental analysis of pollution reduction schemes. *Journal of Environmental Economics and Management,* 65, 469-484. 10.1016/j.jeem.2012.09.001.

CASTAÑEDA, A., DOAN, D., NEWHOUSE, D., NGUYEN, M. C., UEMATSU, H. & AZEVEDO, J. P. 2018. A New Profile of the Global Poor. *World Development,* 101, 250-267. 10.1016/j.worlddev.2017.08.002.

CAVENDISH, W. J. W. D. 2000. Empirical regularities in the poverty-environment relationship of rural households: Evidence from Zimbabwe. 28, 1979-2003.

CEDDIA, M. G. 2019. The impact of income, land, and wealth inequality on agricultural expansion in Latin America. *Proc Natl Acad Sci U S A,* 116, 2527-2532. 10.1073/pnas.1814894116.

CHAUDHURI, A. 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics,* 14, 47-83. 10.1007/s10683-010-9257-1.

CHAVES, L. S. M., FRY, J., MALIK, A., GESCHKE, A., SALLUM, M. A. M. & LENZEN, M. 2020. Global consumption and international trade in deforestation-associated commodities could influence malaria risk. *Nat Commun,* 11, 1258. 10.1038/s41467-020-14954-1.

CHERVIER, C. & COSTEDOAT, S. 2017. Heterogeneous Impact of a Collective Payment for Environmental Services Scheme on Reducing Deforestation in Cambodia. *World Development,* 98, 148-159. 10.1016/j.worlddev.2017.04.014.
CHHATRE, A. & AGRAWAL, A. 2008. Forest commons and local enforcement. *Proc Natl Acad Sci U S A,* 105.
CINYABUGUMA, M., PAGE, T. & PUTTERMAN, L. 2006. Can second-order punishment deter perverse punishment? *Experimental Economics,* 9, 265-279. 10.1007/s10683-006-9127-z.
COASE, R. H. 1960. The problem of social cost. *Classic papers in natural resource economics.* Springer.
COOMES, O. T., TAKASAKI, Y., ABIZAID, C. & BARHAM, B. L. 2010. Floodplain fisheries as natural insurance for the rural poor in tropical forest environments: evidence from Amazonia. *Fisheries Management and Ecology,* 17, 513-521. 10.1111/j.1365-2400.2010.00750.x.
COSTEDOAT, S., KOETSE, M., CORBERA, E. & EZZINE-DE-BLAS, D. 2016. Cash only? Unveiling preferences for a PES contract through a choice experiment in Chiapas, Mexico. *Land Use Policy,* 58, 302-317. 10.1016/j.landusepol.2016.07.023.
COWELL, F. 2011. *Measuring inequality*, Oxford University Press.
CRESPO CUARESMA, J., DANYLO, O., FRITZ, S., MCCALLUM, I., OBERSTEINER, M., SEE, L. & WALSH, B. 2017. Economic Development and Forest Cover: Evidence from Satellite Data. *Sci Rep,* 7, 40678. 10.1038/srep40678.
CSILLIK, O., KUMAR, P., MASCARO, J., O'SHEA, T. & ASNER, G. P. 2019. Monitoring tropical forest carbon stocks and emissions using Planet satellite data. *Sci Rep,* 9, 17831. 10.1038/s41598-019-54386-6.
CULAS, R. J. 2012. REDD and forest transition: Tunneling through the environmental Kuznets curve. *Ecological Economics,* 79, 44-51. 10.1016/j.ecolecon.2012.04.015.
CURTIS, P. G., SLAY, C. M., HARRIS, N. L., TYUKAVINA, A. & HANSEN, M. C. 2018. Classifying drivers of global forest loss. *Science,* 361, 1108-1111.
DAVE, C., ECKEL, C. C., JOHNSON, C. A. & ROJAS, C. 2010. Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty,* 41, 219-243. 10.1007/s11166-010-9103-z.
DAYTON-JOHNSON, J. & BARDHAN, P. 2002. Inequality and Conservation on the Local Commons: A Theoretical Exercise. *The Economic Journal,* 112.
DE GEEST, L. R. & KINGSLEY, D. C. 2019. Endowment heterogeneity, incomplete information & institutional choice in public good experiments. *Journal of Behavioral and Experimental Economics,* 83. 10.1016/j.socec.2019.101478.
DE GEEST, L. R. & KINGSLEY, D. C. 2021. Inequality and Peer Punishment in a Common-Pool Resource Experiment. *Strategic Behavior and the Environment,* 9, 1-26. 10.1561/102.00000099.
DE JANVRY, A., FAFCHAMPS, M. & SADOULET, E. 1991. Peasant Household Behaviour with Missing Markets: Some Paradoxes Explained. *The Economic Journal,* 101, 1400-1417.
DE JANVRY, A. & SADOULET, E. 2016. *Development Economics, Theory and Practice,* New York, Routledge.
DE KONING, F., AGUIÑAGA, M., BRAVO, M., CHIU, M., LASCANO, M., LOZADA, T. & SUAREZ, L. 2011. Bridging the gap between forest conservation and poverty alleviation: the Ecuadorian Socio Bosque program. *Environmental Science & Policy,* 14, 531-542. 10.1016/j.envsci.2011.04.007.
DE MELO, G. & PIAGGIO, M. 2015. The perils of peer punishment: Evidence from a common pool resource framed field experiment. *Ecological Economics,* 120, 376-393. 10.1016/j.ecolecon.2015.05.011.
DOKKEN, T., CAPLOW, S., ANGELSEN, A. & SUNDERLIN, W. 2014. Tenure Issues in REDD+ Pilot Project Sites in Tanzania. *Forests,* 5, 234-255. 10.3390/f5020234.

DUCHELLE, A. E., DE SASSI, C., JAGGER, P., CROMBERG, M., LARSON, A. M., SUNDERLIN, W. D., ATMADJA, S. S., RESOSUDARMO, I. A. P., PRATAMA, C. D. J. E. & SOCIETY 2017. Balancing carrots and sticks in REDD+ implications for social safeguards. 22.
DUCHELLE, A. E., SIMONET, G., SUNDERLIN, W. D. & WUNDER, S. 2018. What is REDD+ achieving on the ground? *Current Opinion in Environmental Sustainability,* 32**,** 134-140. 10.1016/j.cosust.2018.07.001.
EDWARDS, D. P., SOCOLAR, J. B., MILLS, S. C., BURIVALOVA, Z., KOH, L. P. & WILCOVE, D. S. 2019. Conservation of Tropical Forests in the Anthropocene. *Curr Biol,* 29**,** R1008-R1020. 10.1016/j.cub.2019.08.026.
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**,** 131-177. 10.1561/101.00000076.
ENGEL, S., PAGIOLA, S. & WUNDER, S. 2008. Designing payments for environmental services in theory and practice: An overview of the issues. *Ecological Economics,* 65**,** 663-674. 10.1016/j.ecolecon.2008.03.011.
ENGELMANN, D. & NIKIFORAKIS, N. 2015. In the long-run we are all dead: on the benefits of peer punishment in rich environments. *Social Choice and Welfare,* 45**,** 561-577. 10.1007/s00355-015-0884-5.
ERIKSSON, K., STRIMLING, P., ANDERSSON, P. A., AVEYARD, M., BRAUER, M., GRITSKOV, V., KIYONARI, T., KUHLMAN, D. M., MAITNER, A. T., MANESI, Z., MOLHO, C., PEPERKOORN, L. S., RIZWAN, M., STIVERS, A. W., TIAN, Q., VAN LANGE, P. A. M., VARTANOVA, I., WU, J. & YAMAGISHI, T. 2017. Cultural Universals and Cultural Differences in Meta-Norms about Peer Punishment. *Management and Organization Review,* 13**,** 851-870. 10.1017/mor.2017.42.
ERIKSSON, K., STRIMLING, P., GELFAND, M., WU, J., ABERNATHY, J., AKOTIA, C. S., ALDASHEV, A., ANDERSSON, P. A., ANDRIGHETTO, G. & ANUM, A. J. N. C. 2021. Perceptions of the appropriate response to norm violation in 57 societies. *Nature Communications,* 12**,** 1-11.
FALK, A., FEHR, E. & FISCHBACHER, U. 2005. Driving forces behind informal sanctions. *Econometrica,* 73.
FAO 2020. The state of the world's forests 2020: forests, biodiversity and people. Rome: FAO
FARIA, W. R. & ALMEIDA, A. N. 2016. Relationship between openness to trade and deforestation: Empirical evidence from the Brazilian Amazon. *Ecological Economics,* 121**,** 85-97. 10.1016/j.ecolecon.2015.11.014.
FEENSTRA, R. C., INKLAAR, R. & TIMMER, M. P. 2015. The Next Generation of the Penn World Table. *American Economic Review,* 105**,** 3150-3182. 10.1257/aer.20130954.
FEHR, E. & FISCHBACHER, U. 2004. Third-party punishment and social norms. *Evolution and Human Behavior,* 25**,** 63-87. 10.1016/s1090-5138(04)00005-4.
FEHR, E. & GACHTER, S. 2000. Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives,* 14.
FEHR, E., GLÄTZLE-RÜTZLER, D. & SUTTER, M. 2013. The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review,* 64**,** 369-383. 10.1016/j.euroecorev.2013.09.006.
FEHR, E. & GÄCHTER, S. 2000. Cooperation and Punishment in Public Goods Experiments. *American Economic Review,* 90.
FEHR, E. & GÄCHTER, S. 2002. Altruistic Punishment in Humans. *Nature,* 415**,** 137-139. https://doi.org/10.1038/415137a.
FEHR, E. & SCHMIDT, K. 1999. Theory of Fairness, Competition and Cooperation. *The Quaterly Journal of Economics*.
FERRARO, P. J. 2001. Global Habitat Protection: Limitations of Development Interventions and a Role for Conservation Performance Payments. *Conservation Biology,* 15.

FERRARO, P. J. & HANAUER, M. M. 2014. Advances in Measuring the Environmental and Social Impacts of Environmental Programs. *Annual Review of Environment and Resources,* 39**,** 495-517. 10.1146/annurev-environ-101813-013230.

FERRARO, P. J. & HANAUER, M. M. 2015. Through what mechanisms do protected areas affect environmental and social outcomes? *Philos Trans R Soc Lond B Biol Sci,* 370. 10.1098/rstb.2014.0267.

FINKBEINER, E. M., MICHELI, F., SAENZ-ARROYO, A., VAZQUEZ-VERA, L., PERAFAN, C. A. & CÁRDENAS, J. C. 2018. Local response to global uncertainty: Insights from experimental economics in small-scale fisheries. *Global Environmental Change,* 48**,** 151-157. 10.1016/j.gloenvcha.2017.11.010.

FOSTER, A. D. & ROSENWEIG, M. R. 2003. Economic Growth and the Rise of Forests. *Quaterly Journal of Economics,* 118**,** 601-637.

FRIEDMAN, R. S., LAW, E. A., BENNETT, N. J., IVES, C. D., THORN, J. P. R. & WILSON, K. A. 2018. How just and just how? A systematic review of social equity in conservation research. *Environmental Research Letters,* 13. 10.1088/1748-9326/aabcde.

GACHTER, S. & HERRMANN, B. 2009. Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philos Trans R Soc Lond B Biol Sci,* 364**,** 791-806. 10.1098/rstb.2008.0275.

GALIZZI, M. M. & NAVARRO-MARTÍNEZ, D. J. M. S. 2019. On the external validity of social preference games: a systematic lab-field study. *Management Science,* 65**,** 976-1002.

GANGADHARAN, L., NIKIFORAKIS, N. & VILLEVAL, M. C. 2017. Normative conflict and the limits of self-governance in heterogeneous populations. *European Economic Review,* 100**,** 143-156. 10.1016/j.euroecorev.2017.07.013.

GATISO, T. T., VOLLAN, B., VIMAL, R. & KÜHL, H. S. 2018. If Possible, Incentivize Individuals Not Groups: Evidence from Lab-in-the-Field Experiments on Forest Conservation in Rural Uganda. *Conservation Letters,* 11. 10.1111/conl.12387.

GATTI, L. V., BASSO, L. S., MILLER, J. B., GLOOR, M., GATTI DOMINGUES, L., CASSOL, H. L. G., TEJADA, G., ARAGAO, L., NOBRE, C., PETERS, W., MARANI, L., ARAI, E., SANCHES, A. H., CORREA, S. M., ANDERSON, L., VON RANDOW, C., CORREIA, C. S. C., CRISPIM, S. P. & NEVES, R. A. L. 2021. Amazonia as a carbon source linked to deforestation and climate change. *Nature,* 595**,** 388-393. 10.1038/s41586-021-03629-6.

GEIST, H. J. & LAMBIN, E. F. 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation. *BioScience* 52.

GELCICH, S., GUZMAN, R., RODRÍGUEZ-SICKERT, C., CASTILLA, J. C. & CÁRDENAS, J. C. 2013. Exploring External Validity of Common Pool Resource Experiments: Insights from Artisanal Benthic Fisheries in Chile. *Ecology and Society,* 18. 10.5751/es-05598-180302.

GIBB, R., REDDING, D. W., CHIN, K. Q., DONNELLY, C. A., BLACKBURN, T. M., NEWBOLD, T. & JONES, K. E. 2020. Zoonotic host diversity increases in human-dominated ecosystems. *Nature,* 584**,** 398-402. 10.1038/s41586-020-2562-8.

GIBSON, L., LEE, T. M., KOH, L. P., BROOK, B. W., GARDNER, T. A., BARLOW, J., PERES, C. A., BRADSHAW, C. J., LAURANCE, W. F., LOVEJOY, T. E. & SODHI, N. S. 2011. Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature,* 478**,** 378-81. 10.1038/nature10425.

GIUDICE, R. & BÖRNER, J. 2021. Benefits and costs of incentive-based forest conservation in the Peruvian Amazon. *Forest Policy and Economics,* 131. 10.1016/j.forpol.2021.102559.

GIUDICE, R., BÖRNER, J., WUNDER, S. & CISNEROS, E. 2019. Selection biases and spillovers from collective conservation incentives in the Peruvian Amazon. *Environmental Research Letters,* 14. 10.1088/1748-9326/aafc83.

GOLLIN, D., LAGAKOS, D. & WAUGH, M. E. 2014. The Agricultural Productivity Gap. *The Quarterly Journal of Economics,* 129**,** 939-993. 10.1093/qje/qjt056.

GRECHENIG, K., NICKLISCH, A. & THÖNI, C. 2010. Punishment Despite Reasonable Doubt— A Public Goods Experiment with Sanctions Under Uncertainty. *Journal of Empirical Legal Studies,* 7**,** 847–867.

GRISCOM, B. W., ADAMS, J., ELLIS, P. W., HOUGHTON, R. A., LOMAX, G., MITEVA, D. A., SCHLESINGER, W. H., SHOCH, D., SIIKAMAKI, J. V., SMITH, P., WOODBURY, P., ZGANJAR, C., BLACKMAN, A., CAMPARI, J., CONANT, R. T., DELGADO, C., ELIAS, P., GOPALAKRISHNA, T., HAMSIK, M. R., HERRERO, M., KIESECKER, J., LANDIS, E., LAESTADIUS, L., LEAVITT, S. M., MINNEMEYER, S., POLASKY, S., POTAPOV, P., PUTZ, F. E., SANDERMAN, J., SILVIUS, M., WOLLENBERG, E. & FARGIONE, J. 2017. Natural climate solutions. *Proc Natl Acad Sci U S A,* 114**,** 11645-11650. 10.1073/pnas.1710465114.

GROOM, B., GROSJEAN, P., KONTOLEON, A., SWANSON, T. & ZHANG, S. 2009. Relaxing rural constraints: a 'win-win' policy for poverty and environment in China? *Oxford Economic Papers,* 62**,** 132-156. 10.1093/oep/gpp021.

GROOM, B. & PALMER, C. 2010. Cost-effective provision of environmental services: the role of relaxing market constraints. *Environment and Development Economics,* 15**,** 219-240. 10.1017/s1355770x09990167.

GÄCHTER, S. & HERRMANN, B. 2011. The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. *European Economic Review,* 55**,** 193-210. 10.1016/j.euroecorev.2010.04.003.

HACKETT, S., SCHLAGER, E. & WALKER, J. M. 1994. The role of communication in resolving the commons dilemmas: experimental evidence with heterogenous appropriators. *Journal of Environmental Economics and Management,* 27**,** 99-126.

HANDBERG, Ø. N. & ANGELSEN, A. 2015. Experimental tests of tropical forest conservation measures. *Journal of Economic Behavior & Organization,* 118**,** 346-359. 10.1016/j.jebo.2015.03.007.

HANDBERG, Ø. N. & ANGELSEN, A. 2019. Pay little, get little; pay more, get a little more: A framed forest experiment in Tanzania. *Ecological Economics,* 156**,** 454-467. 10.1016/j.ecolecon.2016.09.025.

HANSEN, M. C., POTAPOV, P. V., MOORE, R., HANCHER, M., TURUBANOVA, S. A., TYUKAVINA, A., THAU, D., STEHMAN, S. V., GOETZ, S. J. & LOVELAND, T. R. J. S. 2013. High-resolution global maps of 21st-century forest cover change. 342**,** 850-853.

HARDIN, G. 1968. Tragedy of the commons. *Science,* 162**,** 1443-1448.

HARDING, T., HERZBERG, J. & KURALBAYEVA, K. 2021. Commodity prices and robust environmental regulation: Evidence from deforestation in Brazil. *Journal of Environmental Economics and Management,* 108. 10.1016/j.jeem.2021.102452.

HARRIS, N. L., GIBBS, D. A., BACCINI, A., BIRDSEY, R. A., DE BRUIN, S., FARINA, M., FATOYINBO, L., HANSEN, M. C., HEROLD, M., HOUGHTON, R. A., POTAPOV, P. V., SUAREZ, D. R., ROMAN-CUESTA, R. M., SAATCHI, S. S., SLAY, C. M., TURUBANOVA, S. A. & TYUKAVINA, A. 2021. Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change,* 11**,** 234-240. 10.1038/s41558-020-00976-6.

HARRISON, G. & LIST, J. A. 2004. Field Experiments. *Journal of Economic Literature,* 42**,** 1009-1055.

HARSTAD, B. 2016. The market for conservation and other hostages. *Journal of Economic Theory,* 166**,** 124-151. 10.1016/j.jet.2016.07.003.

HARTWICK, J. M., VAN LONG, N. & TIAN, H. 2001. Deforestation and Development in a Small Open Economy. *Journal of Environmental Economics and Management,* 41**,** 235-251. 10.1006/jeem.1998.1136.

HAUSER, O. P., HILBE, C., CHATTERJEE, K. & NOWAK, M. A. 2019. Social dilemmas among unequals. *Nature,* 572**,** 524-527. 10.1038/s41586-019-1488-5.

HAYES, T., GRILLOS, T., BREMER, L. L., MURTINHO, F. & SHAPIRO, E. 2019. Collective PES: More than the sum of individual incentives. *Environmental Science & Policy,* 102**,** 1-8. 10.1016/j.envsci.2019.09.010.

HAYES, T. & MURTINHO, F. 2018. Communal governance, equity and payment for ecosystem services. *Land Use Policy,* 79**,** 123-136. 10.1016/j.landusepol.2018.08.001.
HAYES, T., MURTINHO, F. & WOLFF, H. 2017. The Impact of Payments for Environmental Services on Communal Lands: An Analysis of the Factors Driving Household Land-Use Behavior in Ecuador. *World Development,* 93**,** 427-446. 10.1016/j.worlddev.2017.01.003.
HAYES, T., MURTINHO, F., WOLFF, H., LÓPEZ-SANDOVAL, M. F. & SALAZAR, J. 2021. Effectiveness of payment for ecosystem services after loss and uncertainty of compensation. *Nature Sustainability*. 10.1038/s41893-021-00804-5.
HENNINGSEN, A. & HENNING, C. H. C. A. 2009. Imposing regional monotonicity on translog stochastic production frontiers with a simple three-step procedure. *Journal of Productivity Analysis,* 32**,** 217-229. 10.1007/s11123-009-0142-x.
HENRICH, J., ENSMINGER, J., MCELREATH, R., BARR, A., BARRETT, C., BOLYANATZ, A., CARDENAS, J. C., GURVEN, M., GWAKO, E., HENRICH, N., LESOROGOL, C., MARLOWE, F., TRACER, D. & ZIKER, J. 2010a. Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. *Science,* 327**,** 1480-1484.
HENRICH, J., HEINE, S. J. & NORENZAYAN, A. J. N. 2010b. Most people are not WEIRD. 466**,** 29-29.
HENRICH, J., MCELREATH, R., BARR, A., ENSMINGER, J., BARRETT, C., BOLYANATZ, A., CARDENAS, J. C., GURVEN, M., GWAKO, E. & HENRICH, N. J. S. 2006. Costly punishment across human societies. 312**,** 1767-1770.
HERRMANN, B., THÖNI, C. & GÄCHTER, S. 2008. Antisocial Punishment across Societies. *Science,* 310. 10.1029/2003jb002610.
HIRANO, K., IMBENS, G. W. & RIDDER, G. J. E. 2003. Efficient estimation of average treatment effects using the estimated propensity score. 71**,** 1161-1189.
HOLT, C. A. 2007. *Markets, games, & strategic behavior*, Pearson Addison Wesley Boston.
HOLT, C. A. & LAURY, S. K. 2002. Risk aversion and incentives effects. *American Economic Review,* 92**,** 1644-1655.
HOSONUMA, N., HEROLD, M., DE SY, V., DE FRIES, R. S., BROCKHAUS, M., VERCHOT, L., ANGELSEN, A. & ROMIJN, E. 2012. An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters,* 7. 10.1088/1748-9326/7/4/044009.
ICKOWITZ, A., SILLS, E. & DE SASSI, C. 2017. Estimating Smallholder Opportunity Costs of REDD+: A Pantropical Analysis from Households to Carbon and Back. *World Development,* 95**,** 15-26. 10.1016/j.worlddev.2017.02.022.
IPCC 2019. Climate Change and Land IPCC Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems IPCC,
IZQUIERDO-TORT, S., CORBERA, E., BARCEINAS CRUZ, A., NAIME, J., ANGÉLICA VÁZQUEZ-CISNEROS, P., CARABIAS LILLO, J., CASTRO-TOVAR, E., ORTIZ ROSAS, F., RUBIO, N., TORRES KNOOP, L. & DUPRAS, J. 2021. Local responses to design changes in payments for ecosystem services in Chiapas, Mexico. *Ecosystem Services,* 50. 10.1016/j.ecoser.2021.101305.
JACK, B. K. & JAYACHANDRAN, S. 2019. Self-selection into payments for ecosystem services programs. *Proc Natl Acad Sci U S A,* 116**,** 5326-5333. 10.1073/pnas.1802868115.
JAGGER, P., LUCKERT, M. K., BANANA, A. & BAHATI, J. 2012. Asking Questions to Understand Rural Livelihoods: Comparing Disaggregated vs. Aggregated Approaches to Household Livelihood Questionnaires. *World Development,* 40**,** 1810-1823. 10.1016/j.worlddev.2012.04.030.
JAGGER, P., SELLERS, S., KITTNER, N., DAS, I. & BUSH, G. K. 2018. Looking for medium-term conservation and development impacts of community management agreements in Uganda's Rwenzori Mountains National Park. *Ecol Econ,* 152**,** 199-206. 10.1016/j.ecolecon.2018.06.006.

JAYACHANDRAN, S. 2013. Liquidity Constraints and Deforestation: The Limitations of Payments for Ecosystem Services. *American Economic Review,* 103**,** 309-313. 10.1257/aer.103.3.309.
JERVEN, M. 2013. *Poor numbers*, Cornell University Press.
JONES, K. W., ETCHART, N., HOLLAND, M., NAUGHTON-TREVES, L. & ARRIAGADA, R. 2020. The impact of paying for forest conservation on perceived tenure security in Ecuador. *Conservation Letters,* 13. 10.1111/conl.12710.
JOPPA, L. & PFAFF, A. 2010. Reassessing the forest impacts of protection: the challenge of nonrandom location and a corrective method. *PLOS One,* 1185**,** 135-149.
JOPPA, L. N. & PFAFF, A. 2011. Global protected area impacts. *Proc Biol Sci,* 278**,** 1633-8. 10.1098/rspb.2010.1713.
KACZAN, D., PFAFF, A., RODRIGUEZ, L. & SHAPIRO-GARZA, E. 2017. Increasing the impact of collective incentives in payments for ecosystem services. *Journal of Environmental Economics and Management,* 86**,** 48-67. 10.1016/j.jeem.2017.06.007.
KAIMOWITZ, D. & ANGELSEN, A. 1998. *Economic models of tropical deforestation: a review,* Bogor, Indonesia, CIFOR.
KARSENTY, A. & ONGOLO, S. 2012. Can "fragile states" decide to reduce their deforestation? The inappropriate use of the theory of incentives with respect to the REDD mechanism. *Forest Policy and Economics,* 18**,** 38-45. 10.1016/j.forpol.2011.05.006.
KAUFMANN, D. 2007. *The worldwide governance indicators project: answering the critics*, World Bank Publications.
KAZUNGU, M., ZHUNUSOVA, E., KABWE, G. & GÜNTER, S. 2021. Household-Level Determinants of Participation in Forest Support Programmes in the Miombo Landscapes, Zambia. *Sustainability,* 13. 10.3390/su13052713.
KEENAN, R. J., REAMS, G. A., ACHARD, F., DE FREITAS, J. V., GRAINGER, A. & LINDQUIST, E. 2015. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management,* 352**,** 9-20. 10.1016/j.foreco.2015.06.014.
KENT, D. 2020. Comparing alternative estimation methods of a public goods game. *Journal of the Economic Science Association,* 6**,** 156-167. 10.1007/s40881-020-00092-3.
KERR, J., VARDHAN, M. & JINDAL, R. 2012. Prosocial behavior and incentives: Evidence from field experiments in rural Mexico and Tanzania. *Ecological Economics,* 73**,** 220-227. 10.1016/j.ecolecon.2011.10.031.
KERR, S. C. J. R. O. E. E. & POLICY 2013. The economics of international policy agreements to reduce emissions from deforestation and degradation. 7**,** 47-66.
KESSLER, J. B. & VESTERLUND, L. 2015. The external validity of laboratory experiments: Qualitative rather than quantitative effects.
KHANDKER, S., B. KOOLWAL, G. & SAMAD, H. 2009. *Handbook on impact evaluation: quantitative methods and practices*, The World Bank.
KINGSLEY, D. C. 2016. Endowment heterogeneity and peer punishment in a public good experiment: Cooperation and normative conflict. *Journal of Behavioral and Experimental Economics,* 60**,** 49-61. 10.1016/j.socec.2015.12.002.
KIRKPATRICK, C. & BARRIENTOS, A. 2004. The lewis model after 50 years. *The Manchester School,* 72.
KOCH, C., NIKIFORAKIS, N. & NOUSSAIR, C. N. 2021. Covenants before the swords: The limits to efficient cooperation in heterogeneous groups. *Journal of Economic Behavior & Organization,* 188**,** 307-321. 10.1016/j.jebo.2021.05.003.
KOOP, G. & TOLE, L. 1999. Is there an environmental Kuznets curve for deforestation? *Journal of Development Economics,* 58**,** 231-244.
KOOP, G. & TOLE, L. 2001. Deforestation, distribution and development. *Global Environmental Change,* 11**,** 193-202.
KOSFELD, M., OKADA, A. & RIEDL, A. 2009. Institution Formation in Public Goods Games. *American Economic Review,* 99**,** 1335-1355. 10.1257/aer.99.4.1335.

KOSFELD, M. & RUSTAGI, D. 2015. Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia. *American Economic Review,* 105**,** 747-783. 10.1257/aer.20120700.
KOTCHEN, M. J. & SEGERSON, K. 2019. On the use of group performance and rights for environmental protection and resource management. *Proc Natl Acad Sci U S A,* 116**,** 5285-5292. 10.1073/pnas.1802881115.
KRAUSE, T., COLLEN, W. & NICHOLAS, K. A. 2013. Evaluating Safeguards in a Conservation Incentive Program: Participation, Consent, and Benefit Sharing in Indigenous Communities of the Ecuadorian Amazon. *Ecology and Society,* 18. 10.5751/es-05733-180401.
KRUEGER, A. O., SCHIFF, M. & VALDÉS, A. J. T. W. B. E. R. 1988. Agricultural incentives in developing countries: Measuring the effect of sectoral and economywide policies. 2**,** 255-271.
KUBE, S. & TRAXLER, C. 2011. The interaction of legal and social norm enforcement. *Journal of Public Economic Theory,* 13**,** 639–660.
KÖTHKE, M., LEISCHNER, B. & ELSASSER, P. 2013. Uniform global deforestation patterns — An empirical analysis. *Forest Policy and Economics,* 28**,** 23-37. 10.1016/j.forpol.2013.01.001.
LAGAKOS, D. & WAUGH, M. E. 2013. Selection, Agriculture, and Cross-country productivity differences. *American Economic Review,* 103**,** 948-980. 10.1257/aer.
LAMBIN, E. F. & MEYFROIDT, P. 2010. Land use transitions: Socio-ecological feedback versus socio-economic change. *Land Use Policy,* 27**,** 108-118. 10.1016/j.landusepol.2009.09.003.
LE VELLY, G. & DUTILLY, C. 2016. Evaluating Payments for Environmental Services: Methodological Challenges. *PLoS One,* 11**,** e0149374. 10.1371/journal.pone.0149374.
LEBLOIS, A., DAMETTE, O. & WOLFERSBERGER, J. 2017. What has Driven Deforestation in Developing Countries Since the 2000s? Evidence from New Remote-Sensing Data. *World Development,* 92**,** 82-102. 10.1016/j.worlddev.2016.11.012.
LESZCZENSKY, L. & WOLBRING, T. 2019. How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study. *Sociological Methods & Research*. 10.1177/0049124119882473.
LEVITT, S. D. & LIST, J. A. 2007. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives,* 21**,** 153-174.
LEWIS, A. 1954. Economic development with unlimited supplies of labour.
LIU, Z. & KONTOLEON, A. 2018. Meta-Analysis of Livelihood Impacts of Payments for Environmental Services Programmes in Developing Countries. *Ecological Economics,* 149**,** 48-61. 10.1016/j.ecolecon.2018.02.008.
LIU, Z. & LAN, J. 2015. The Sloping Land Conversion Program in China: Effect on the Livelihood Diversification of Rural Households. *World Development,* 70**,** 147-161. 10.1016/j.worlddev.2015.01.004.
LLISO, B., PASCUAL, U. & ENGEL, S. 2021. On the role of social equity in payments for ecosystem services in Latin America: A practitioner perspective. *Ecological Economics,* 182. 10.1016/j.ecolecon.2020.106928.
LOFT, L., GEHRIG, S., SALK, C. & ROMMEL, J. 2020. Fair payments for effective environmental conservation. *Proc Natl Acad Sci U S A,* 117**,** 14094-14101. 10.1073/pnas.1919783117.
LOFT, L., LE, D. N., PHAM, T. T., YANG, A. L., TJAJADI, J. S. & WONG, G. Y. 2017. Whose Equity Matters? National to Local Equity Perceptions in Vietnam's Payments for Forest Ecosystem Services Scheme. *Ecological Economics,* 135**,** 164-175. 10.1016/j.ecolecon.2017.01.016.

LOPEZ, M. C., MURPHY, J. J., SPRAGGON, J. M. & STRANLUND, J. K. 2012. Comparing the Effectiveness of Regulation and Pro-Social Emotions to Enhance Cooperation: Experimental Evidence from Fishing Communities in Colombia. *Economic Inquiry,* 50**,** 131-142. 10.1111/j.1465-7295.2010.00344.x.

LOPEZ, M. C., MURPHY, J. J., SPRAGGON, J. M. & STRANLUND, J. K. 2013. Does government regulation complement existing community efforts to support cooperation? Evidence from field experiments in Colombia. *Handbook on Experimental Economics and the Environment.* Edward Elgar Publishing.

LUND, J. F., SHACKLETON, S. & LUCKERT, M. 2011. Getting quality data. *In:* ANGELSEN, A., LARSEN, H. O. & OLSEN, C. S. (eds.) *Measuring livelihoods environmental dependence: Methods for research fieldwork.* London: Routledge.

MALANI, A. & REIF, J. 2015. Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform. *Journal of Public Economics,* 124**,** 1-17. 10.1016/j.jpubeco.2015.01.001.

MARTIN, A., GROSS-CAMP, N., KEBEDE, B. & MCGUIRE, S. 2014. Measuring effectiveness, efficiency and equity in an experimental Payments for Ecosystem Services trial. *Global Environmental Change,* 28**,** 216-226. 10.1016/j.gloenvcha.2014.07.003.

MARTINANGELI, A. F. M. & MARTINSSON, P. 2020. We, the rich: Inequality, identity and cooperation. *Journal of Economic Behavior & Organization,* 178**,** 249-266. 10.1016/j.jebo.2020.07.013.

MASCLET, D., NOUSSAIR, C., TUCKER, S. & VILLEVAL, M. C. 2003. Monetary and Non monetary Punishment in the Voluntary Contributions Mechanism. *The American Economic Review*.

MASSARELLA, K., SALLU, S. M., ENSOR, J. E. & MARCHANT, R. 2018. REDD+, hype, hope and disappointment: The dynamics of expectations in conservation and development pilot projects. *World Development,* 109**,** 375-385. 10.1016/j.worlddev.2018.05.006.

MATHER, A. S. J. A. 1992. The forest transition. 367-379.

MEYFROIDT, P. 2015. Approaches and terminology for causal analysis in land systems science. *Journal of Land Use Science,* 11**,** 501-522. 10.1080/1747423x.2015.1117530.

MEYFROIDT, P., RUDEL, T. K. & LAMBIN, E. F. 2010. Forest transitions, trade, and the global displacement of land use. *Proc Natl Acad Sci U S A,* 107**,** 20917-22. 10.1073/pnas.1014773107.

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. *Ecological Economics,* 120**,** 394-405. 10.1016/j.ecolecon.2015.04.006.

MIN-VENDITTI, A. A., MOORE, G. W. & FLEISCHMAN, F. 2017. What policies improve forest cover? A systematic review of research from Mesoamerica. *Global Environmental Change,* 47**,** 21-27. 10.1016/j.gloenvcha.2017.08.010.

MINAM 2020. Mapa de la cobertura y pérdida de bosque húmedo amazónico al 2019. *In:* AMBIENTE, M. D. M. (ed.). Lima, Peru,

MITCHARD, E. T. A. 2018. The tropical forest carbon cycle and climate change. *Nature,* 559**,** 527-534. 10.1038/s41586-018-0300-2.

MOFFATT, P. G. 2015. *Experimetrics: Econometrics for experimental economics*, Macmillan International Higher Education.

MONTOYA-ZUMAETA, J., ROJAS, E. & WUNDER, S. 2019. Adding rewards to regulation: The impacts of watershed conservation on land cover and household wellbeing in Moyobamba, Peru. *PLoS One,* 14**,** e0225367. 10.1371/journal.pone.0225367.

MONTOYA-ZUMAETA, J. G., WUNDER, S. & TACCONI, L. 2021. Incentive-based conservation in Peru: Assessing the state of six ongoing PES and REDD+ initiatives. *Land Use Policy,* 108. 10.1016/j.landusepol.2021.105514.

MORGAN, S. L. & WINSHIP, C. 2015. *Counterfactuals and causal inference*, Cambridge University Press.

MULLER, J. & ALBERS, H. J. 2004. Enforcement, payments, and development projects near protected areas: how the market setting determines what works where. *Resource and Energy Economics,* 26**,** 185-204. 10.1016/j.reseneeco.2003.11.007.
MUÑOZ-PIÑA, C., GUEVARA, A., TORRES, J. M. & BRAÑA, J. 2008. Paying for the hydrological services of Mexico's forests: Analysis, negotiations and results. *Ecological Economics,* 65**,** 725-736. 10.1016/j.ecolecon.2007.07.031.
MURTINHO, F. & HAYES, T. 2017. Communal Participation in Payment for Environmental Services (PES): Unpacking the Collective Decision to Enroll. *Environmental Management,* 59**,** 939-955. 10.1007/s00267-017-0838-z.
NAIDOO, R., WEAVER, L. C., DIGGLE, R. W., MATONGO, G., STUART-HILL, G. & THOULESS, C. 2016. Complementary benefits of tourism and hunting to communal conservancies in Namibia. *Conserv Biol,* 30**,** 628-38. 10.1111/cobi.12643.
NAIDU, S. C. 2009. Heterogeneity and Collective Management: Evidence from Common Forests in Himachal Pradesh, India. *World Development,* 37**,** 676-678. 10.1016/j.worlddev.2008.07.001.
NARLOCH, U., PASCUAL, U. & DRUCKER, A. G. 2012. Collective Action Dynamics under External Rewards: Experimental Insights from Andean Farming Communities. *World Development,* 40**,** 2096-2107. 10.1016/j.worlddev.2012.03.014.
NEWTON, P., KINZER, A. T., MILLER, D. C., OLDEKOP, J. A. & AGRAWAL, A. 2020. The Number and Spatial Distribution of Forest-Proximate People Globally. *One Earth,* 3**,** 363-370. 10.1016/j.oneear.2020.08.016.
NEWTON, P., MILLER, D. C., BYENKYA, M. A. A. & AGRAWAL, A. 2016. Who are forest-dependent people? A taxo nomy to aid livelihood and land use decision-making in forested regions. *Land Use Policy,* 57**,** 388-395. 10.1016/j.landusepol.2016.05.032.
NGOMA, H., HAILU, A. T., KABWE, S. & ANGELSEN, A. 2020. Pay, talk or 'whip' to conserve forests: Framed field experiments in Zambia. *World Development,* 128. 10.1016/j.worlddev.2019.104846.
NGUYEN, T. L., COLLINS, G. S., SPENCE, J., DAURES, J. P., DEVEREAUX, P. J., LANDAIS, P. & LE MANACH, Y. 2017. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. *BMC Med Res Methodol,* 17**,** 78. 10.1186/s12874-017-0338-0.
NIKIFORAKIS, N. 2008. Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics,* 92**,** 91-112. 10.1016/j.jpubeco.2007.04.008.
NIKIFORAKIS, N. & NORMANN, H.-T. 2008. A comparative statics analysis of punishment in public-good experiments. *Experimental Economics,* 11**,** 358-369.
NIKIFORAKIS, N., NOUSSAIR, C. N. & WILKENING, T. 2012. Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics,* 96**,** 797-807. 10.1016/j.jpubeco.2012.05.014.
NOCKUR, L., PFATTHEICHER, S. & KELLER, J. 2021. Different punishment systems in a public goods game with asymmetric endowments. *Journal of Experimental Social Psychology,* 93. 10.1016/j.jesp.2020.104096.
NOON, M. L., GOLDSTEIN, A., LEDEZMA, J. C., ROEHRDANZ, P. R., COOK-PATTON, S. C., SPAWN-LEE, S. A., WRIGHT, T. M., GONZALEZ-ROGLICH, M., HOLE, D. G., ROCKSTRÖM, J. & TURNER, W. R. 2021. Mapping the irrecoverable carbon in Earth's ecosystems. *Nature Sustainability*. 10.1038/s41893-021-00803-6.
NOSENZO, D. & SEFTON, M. 2014. Promoting cooperation: The distribution of reward and punishment power. *In:* VAN LANGE, P. A. M., ROCKENBACH, B. & YAMAGISHI, T. (eds.) *Social dilemmas: New perspectives on reward and punishment* Oxford: Oxford University Press.
NOUSSAIR, C. & TUCKER, S. 2005. Combining Monetary and Social Sanctions to Promote Cooperation. *Economic Inquiry,* 43**,** 649-660. 10.1093/ei/cbi045.
NYDF, A. P. 2021. Taking stock of national climate action for forests. . *In:* FOCUS, C. (ed.).

OLA, O., MENAPACE, L., BENJAMIN, E. & LANG, H. 2019. Determinants of the environmental conservation and poverty alleviation objectives of Payments for Ecosystem Services (PES) programs. *Ecosystem Services,* 35**,** 52-66. 10.1016/j.ecoser.2018.10.011.

OLDEKOP, J. A., SIMS, K. R. E., KARNA, B. K., WHITTINGHAM, M. J. & AGRAWAL, A. 2019. Reductions in deforestation and poverty from decentralized forest management in Nepal. *Nature Sustainability,* 2**,** 421-428. 10.1038/s41893-019-0277-3.

OLIVERO, J., FA, J. E., REAL, R., MARQUEZ, A. L., FARFAN, M. A., VARGAS, J. M., GAVEAU, D., SALIM, M. A., PARK, D., SUTER, J., KING, S., LEENDERTZ, S. A., SHEIL, D. & NASI, R. 2017. Recent loss of closed forests is associated with Ebola virus disease outbreaks. *Sci Rep,* 7**,** 14291. 10.1038/s41598-017-14727-9.

ONES, U. & PUTTERMAN, L. 2007. The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior & Organization,* 62**,** 495-521. 10.1016/j.jebo.2005.04.018.

OSTROM, E. 1990. *Governing the commons: The evolution of institutions for collective action*, Cambridge university press.

OSTROM, E. 1993. Design Principles in Long-Enduring Irrigation Institutions. *Water Resources Research,* 29.

OSTROM, E. 1998. A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997. *American Political Science Review,* 92**,** 1-22. 10.2307/2585925.

OSTROM, E. 2006. The value-added of laboratory experiments for the study of institutions and common-pool resources. *Journal of Economic Behavior & Organization,* 61**,** 149-163. 10.1016/j.jebo.2005.02.008.

OSTROM, E. 2010. Polycentric systems for coping with collective action and global environmental change. *Global Environmental Change,* 20**,** 550-557. 10.1016/j.gloenvcha.2010.07.004.

OSTROM, E., WALKER, J. M. & GARDNER, R. 1992. Covenants with and without a sword: self governance is possible. *American Political Science Review,* 86**,** 404-417.

OZONO, H., KAMIJO, Y. & SHIMIZU, K. 2017. Punishing second-order free riders before first-order free riders: The effect of pool punishment priority on cooperation. *Scientific reports,* 7**,** 1-9.

PAGIOLA, S., ARCENAS, A. & PLATAIS, G. 2005. Can Payments for Environmental Services Help Reduce Poverty? An Exploration of the Issues and the Evidence to Date from Latin America. *World Development,* 33**,** 237-253. 10.1016/j.worlddev.2004.07.011.

PAGIOLA, S., HONEY-ROSES, J. & FREIRE-GONZALEZ, J. 2016. Evaluation of the Permanence of Land Use Change Induced by Payments for Environmental Services in Quindio, Colombia. *PLoS One,* 11**,** e0147829. 10.1371/journal.pone.0147829.

PALMER, C., SOUZA, G. I., LARAY, E., VIANA, V. & HALL, A. 2020. Participatory policies and intrinsic motivation to conserve forest commons. *Nature Sustainability,* 3**,** 620-627. 10.1038/s41893-020-0531-8.

PARCO, J. E., RAPOPORT, A. & STEIN, W. E. J. P. S. 2002. Effects of financial incentives on the breakdown of mutual trust. 13**,** 292-297.

PASCUAL, U., MURADIAN, R., RODRÍGUEZ, L. C. & DURAIAPPAH, A. 2010. Exploring the links between equity and efficiency in payments for environmental services: A conceptual approach. *Ecological Economics,* 69**,** 1237-1244. 10.1016/j.ecolecon.2009.11.004.

PASCUAL, U., PHELPS, J., GARMENDIA, E., BROWN, K., CORBERA, E., MARTIN, A., GOMEZ-BAGGETHUN, E. & MURADIAN, R. 2014. Social Equity Matters in Payments for Ecosystem Services. *BioScience,* 64**,** 1027-1036. 10.1093/biosci/biu146.

PATTANAYAK, S. K., WUNDER, S. & FERRARO, P. J. 2010. Show Me the Money: Do Payments Supply Environmental Services in Developing Countries? *Review of Environmental Economics and Policy,* 4**,** 254-274. 10.1093/reep/req006.

PAYNE, B. K., BROWN-IANNUZZI, J. L. & HANNAY, J. W. 2017. Economic inequality increases risk taking. *Proc Natl Acad Sci U S A,* 114**,** 4643-4648. 10.1073/pnas.1616453114.

PENDRILL, F., PERSSON, U. M., GODAR, J., KASTNER, T., MORAN, D., SCHMIDT, S. & WOOD, R. 2019. Agricultural and forestry trade drives large share of tropical deforestation emissions. *Global Environmental Change,* 56**,** 1-10. 10.1016/j.gloenvcha.2019.03.002.
PEREVOCHTCHIKOVA, M., CASTRO-DÍAZ, R., LANGLE-FLORES, A. & VON THADEN UGALDE, J. J. 2021. A systematic review of scientific publications on the effects of payments for ecosystem services in Latin America, 2000–2020. *Ecosystem Services,* 49. 10.1016/j.ecoser.2021.101270.
PERSHA, L. & ANDERSSON, K. 2014. Elite capture risk and mitigation in decentralized forest governance regimes. *Global Environmental Change,* 24**,** 265-276. 10.1016/j.gloenvcha.2013.12.005.
PFAFF, A. & ROBALINO, J. 2012. Protecting forests, biodiversity, and the climate: predicting policy impact to improve policy choice. *Oxford Review of Economic Policy,* 28**,** 164-179. 10.1093/oxrep/grs012.
PFAFF, A., RODRIGUEZ, L. A. & SHAPIRO-GARZA, E. 2019. Collective Local Payments for ecosystem services: New local PES between groups, sanctions, and prior watershed trust in Mexico. *Water Resources and Economics,* 28. 10.1016/j.wre.2019.01.002.
PFATTHEICHER, S., BÖHM, R. & KESBERG, R. 2018. The Advantage of Democratic Peer Punishment in Sustaining Cooperation within Groups. *Journal of Behavioral Decision Making,* 31**,** 562-571. 10.1002/bdm.2050.
PHAM, T. T., CAMPBELL, B. M., GARNETT, S., ASLIN, H. & HOANG, M. H. J. E. C. 2010. Importance and impacts of intermediary boundary organizations in facilitating payment for environmental services in Vietnam. 37**,** 64-72.
PIGOU, A. 1920. *The Economics of Welfare*.
PIRARD, R., WUNDER, S., DUCHELLE, A. E., PURI, J., ASFAW, S., BULUSU, M., PETIT, H. & VEDOVETO, M. 2019. Effectiveness of Forest Conservation Interventions: An Evidence Gap Map. *IEU LEARNING PAPER* Songdo, South Korea: Green Climate Fund,
PORRO, R., LOPEZ-FELDMAN, A. & VELA-ALVARADO, J. W. 2015. Forest use and agriculture in Ucayali, Peru: Livelihood strategies, poverty and wealth in an Amazon frontier. *Forest Policy and Economics,* 51**,** 47-56. 10.1016/j.forpol.2014.12.001.
PROBST, B., BENYISHAY, A., KONTOLEON, A. & DOS REIS, T. N. P. 2020. Impacts of a large-scale titling initiative on deforestation in the Brazilian Amazon. *Nature Sustainability,* 3**,** 1019-1026. 10.1038/s41893-020-0537-2.
PURI, J., NATH, M., BHATIA, R. & GLEW, L. J. I. I. F. I. E., NEW DELHI 2016. Examining the evidence base for forest conservation interventions, Evidence Gap Map Report 4.
QIU, J., GAME, E. T., TALLIS, H., OLANDER, L. P., GLEW, L., KAGAN, J. S., KALIES, E. L., MICHANOWICZ, D., PHELAN, J., POLASKY, S., REED, J., SILLS, E. O., URBAN, D. & WEAVER, S. K. 2018. Evidence-Based Causal Chains for Linking Health, Development, and Conservation Actions. *Bioscience,* 68**,** 182-193. 10.1093/biosci/bix167.
RABE-HESKETH, S. & SKRONDAL, A. 2008. *Multilevel and longitudinal modeling using Stata*, STATA press.
RAVEN, P. H., GEREAU, R. E., PHILLIPSON, P. B., CHATELAIN, C., JENKINS, C. N. & ULLOA, C. U. J. S. A. 2020. The distribution of biodiversity richness in the tropics. 6**,** eabc6228.
RESTUCCIA, D., YANG, D. T. & ZHU, X. 2008. Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics,* 55**,** 234-250. 10.1016/j.jmoneco.2007.11.006.
REUBEN, E. & RIEDL, A. 2013. Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior,* 77**,** 122-137. 10.1016/j.geb.2012.10.001.

ROBALINO, J., PFAFF, A., SANDOVAL, C. & SANCHEZ-AZOFEIFA, G. A. 2021. Can we increase the impacts from payments for ecosystem services? Impact rose over time in Costa Rica, yet spatial variation indicates more potential. *Forest Policy and Economics,* 132. 10.1016/j.forpol.2021.102577.
ROBINSON, E., ALBERS, H., LOKINA, R. & MESHACK, C. 2016. Allocating Group-Level Payments for Ecosystem Services: Experiences from a REDD+ Pilot in Tanzania. *Resources,* 5. 10.3390/resources5040043.
ROBINSON, E. J. Z., KUMAR, A. M. & ALBERS, H. J. 2010. Protecting Developing Countries' Forests: Enforcement in Theory and Practice. *Journal of Natural Resources Policy Research,* 2**,** 25-38. 10.1080/19390450903350820.
RODRIGUEZ-SICKERT, C., GUZMÁN, R. A. & CÁRDENAS, J. C. 2008. Institutions influence preferences: Evidence from a common pool resource experiment. *Journal of Economic Behavior & Organization,* 67**,** 215-227. 10.1016/j.jebo.2007.06.004.
RODRIGUEZ-WARD, D. & PAREDES, P. 2014. Valuation of Environmental Services in the Managed Forests of Seven Indigenous Communities in Ucayali, Peru. *In:* SILLS, E. O., ATMADJA, S. S., DE SASSI, C., DUCHELLE, A. E., KWEKA, D. L., RESOSUDARMO, I. A. P. & SUNDERLIN, W. D. (eds.) *REDD+ on the ground A case book of subnational initiatives across the globe.* Bogor, Indonesia: Cifor.
RODRIGUEZ, L. A., PFAFF, A. & VELEZ, M. A. 2019. Graduated stringency within collective incentives for group environmental compliance: Building coordination in field-lab experiments with artisanal gold miners in Colombia. *Journal of Environmental Economics and Management,* 98. 10.1016/j.jeem.2019.102276.
ROODMAN, D. J. T. S. J. 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. 9**,** 86-136.
ROSENBAUM, P. R. & RUBIN, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70**,** 41-55.
ROSENBAUM, P. R. & RUBIN, D. B. 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician,* 39**,** 33-39.
ROTH, A. E. 2015. Is Experimental Economics Living Up to Its Promise? . *In:* FRÉCHETTE, G. R. & SCHOTTER, A. (eds.) *Handbook of Experimental Economic Methodology.* New York USA: Oxford University Press.
ROTH, A. E., MURNIGHAN, J. K. & SCHOUMAKER, F. J. T. A. E. R. 1988. The deadline effect in bargaining: Some experimental evidence. 78**,** 806-823.
RRI 2015. Who Owns the World's Land? A global baseline of formally recognized indigenous and community land rights. Washington, DC: , Rights and Resources Initiative,
RUBIN, D. B. 1973. The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies. *Biometrics,* 29**,** 185-203.
RUBIN, D. B. 2001. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services and Outcomes Research Methodology,* 2**,** 169-188. 10.1023/A:1020363010465.
RUDEL, T. K., COOMES, O. T., MORAN, E., ACHARD, F., ANGELSEN, A., XU, J. & LAMBIN, E. 2005. Forest transitions: towards a global understanding of land use change. *Global Environmental Change,* 15**,** 23-31. 10.1016/j.gloenvcha.2004.11.001.
RUSTAGI, D., ENGEL, S. & KOSFELD, M. 2010. Conditional Cooperation and Costly Monitoring Explains Success in Forest Commons Management. *Science,* 330**,** 961-963.
SALZMAN, J., BENNETT, G., CARROLL, N., GOLDSTEIN, A. & JENKINS, M. 2018. The global status and trends of Payments for Ecosystem Services. *Nature Sustainability,* 1**,** 136-144. 10.1038/s41893-018-0033-0.
SAMII, C., LISIECKI, M., KULKARNI, P., PALER, L. & CHAVIS, L. 2014. Effects of Payment for Environmental Services and Decentralized Forest Management on Deforestation and Poverty in Low-and Middle-Income Countries: a Systematic Review. *Campbell Systematic Reviews,* 11**,** 1-40.

SCHRECKENBERG, K., CAMARGO, I., WITHNALL, K., CORRIGAN, C., FRANKS, P., ROE, D., SCHERL, L. M. & RICHARDSON, V. 2010. Social assessment of conservation initiatives. *Social Assessment of Conservation Initiatives: A review of rapid methodologies, Natural Resource Issues No. 22.*: IIED London.,

SCHRÖTER, B., HAUCK, J., HACKENBERG, I. & MATZDORF, B. 2018. Bringing transparency into the process: Social network analysis as a tool to support the participatory design and implementation process of Payments for Ecosystem Services. *Ecosystem Services,* 34**,** 206-217. 10.1016/j.ecoser.2018.03.007.

SCOONES, I. 2009. Livelihoods perspectives and rural development. *The Journal of Peasant Studies,* 36**,** 171-196. 10.1080/03066150902820503.

SCOONES, I., MELNYK, M. & PRETTY, J. N. 1992. The hidden harvest: wild foods and agricultural systems; a literature review and annotated bibliography.

SHANEE, N. & SHANEE, S. 2016. Land Trafficking, Migration, and Conservation in the "No-Man's Land" of Northeastern Peru. *Tropical Conservation Science,* 9. 10.1177/1940082916682957.

SHEIL, D. 2018. Forests, atmospheric water and an uncertain future: the new biology of the global water cycle. *Forest Ecosystems,* 5. 10.1186/s40663-018-0138-y.

SHREEDHAR, G., TAVONI, A. & MARCHIORI, C. 2020. Monitoring and punishment networks in an experimental common pool resource dilemma. *Environment and Development Economics,* 25**,** 66-94. 10.1017/s1355770x19000457.

SILLS, E. O., ATMADJA, S. S., DE SASSI, C., DUCHELLE, A. E., KWEKA, D. L., RESOSUDARMO, I. A. P. & SUNDERLIN, W. D. 2014. *REDD+ on the ground: A case book of subnational initiatives across the globe*, Cifor.

SILLS, E. O., DE SASSI, C., JAGGER, P., LAWLOR, K., MITEVA, D. A., PATTANAYAK, S. K. & SUNDERLIN, W. D. 2017. Building the evidence base for REDD+: Study design and methods for evaluating the impacts of conservation interventions on local well-being. *Global Environmental Change,* 43**,** 148-160. 10.1016/j.gloenvcha.2017.02.002.

SILLS, E. O. & JONES, K. 2018. Causal inference in environmental conservation: The role of institutions. *Handbook of environmental economics.* Elsevier.

SIMONET, G. & SEYLLER, C. 2015. Id-recco, a new collaborative work tool to improve knowledge on redd+ projects: sources, methodology and data.

SIMONET, G., SUBERVIE, J., EZZINE-DE-BLAS, D., CROMBERG, M. & DUCHELLE, A. E. 2018. Effectiveness of a REDD+ Project in Reducing Deforestation in the Brazilian Amazon. *American Journal of Agricultural Economics,* 101**,** 211-229. 10.1093/ajae/aay028.

SINGH, I., SQUIRE, L. & STRAUSS, J. 1986. A survey of agricultural household models: Recent findings and policy implications. *The World Bank Economic Review,* 1**,** 149-179.

SKRONDAL, A. & RABE-HESKETH, S. 2004. *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*, Crc Press.

SLOUGH, T., KOPAS, J. & URPELAINEN, J. 2021a. Satellite-based deforestation alerts with training and incentives for patrolling facilitate community monitoring in the Peruvian Amazon. *Proc Natl Acad Sci U S A,* 118. 10.1073/pnas.2015171118.

SLOUGH, T., RUBENSON, D., LEVY, R., ALPIZAR RODRIGUEZ, F., BERNEDO DEL CARPIO, M., BUNTAINE, M. T., CHRISTENSEN, D., COOPERMAN, A., EISENBARTH, S., FERRARO, P. J., GRAHAM, L., HARTMAN, A. C., KOPAS, J., MCLARTY, S., RIGTERINK, A. S., SAMII, C., SEIM, B., URPELAINEN, J. & ZHANG, B. 2021b. Adoption of community monitoring improves common pool resource management across contexts. *Proc Natl Acad Sci U S A,* 118. 10.1073/pnas.2015367118.

SOLIS, D., CRONKLETON, P., SILLS, E. O., RODRIGUEZ-WARD, D. & DUCHELLE, A. E. 2021. Evaluating the Impact of REDD+ Interventions on Household Forest Revenue in Peru. *Frontiers in Forests and Global Change,* 4. 10.3389/ffgc.2021.624724.

SONG, X. P., HANSEN, M. C., STEHMAN, S. V., POTAPOV, P. V., TYUKAVINA, A., VERMOTE, E. F. & TOWNSHEND, J. R. 2018. Global land change from 1982 to 2016. *Nature,* 560**,** 639-643. 10.1038/s41586-018-0411-9.

SPRAGGON, J. M., SOBARZO, L. A. V. & STRANLUND, J. K. 2015. A note on stochastic public revelation and voluntary contributions to public goods. *Economics Letters,* 126**,** 144-146. 10.1016/j.econlet.2014.12.009.
STEFFEN, W., RICHARDSON, K., ROCKSTRÖM, J., CORNELL, S. E., FETZER, I., BENNETT, E. M., BIGGS, R., CARPENTER, S. R., DE VRIES, W. & DE WIT, C. A. J. S. 2015. Planetary boundaries: Guiding human development on a changing planet. 347.
STOCK, J. H. & WATSON, M. W. J. A. 2015. Introduction to econometrics (3rd updated edition). 3.
STUART, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Stat Sci,* 25**,** 1-21. 10.1214/09-STS313.
STURGIS, P. & SMITH, P. 2010. Assessing the Validity of Generalized Trust Questions: What Kind of Trust are we Measuring? *International Journal of Public Opinion Research,* 22**,** 74-92. 10.1093/ijpor/edq003.
SUNDERLIN, W. D., DE SASSI, C., SILLS, E. O., DUCHELLE, A. E., LARSON, A. M., RESOSUDARMO, I. A. P., AWONO, A., KWEKA, D. L. & HUYNH, T. B. 2018. Creating an appropriate tenure foundation for REDD+: The record to date and prospects for the future. *World Development,* 106**,** 376-392. 10.1016/j.worlddev.2018.01.010.
SUNDERLIN, W. D., LARSON, A. M., DUCHELLE, A. E., SILLS, E. O., LUTTRELL, C., JAGGER, P., PATTANAYAK, S. K., CRONKLETON, P., EKAPUTRI, A. D. & DE SASSI, C. 2016. *Technical guidelines for research on REDD+ subnational initiatives*, CIFOR.
SUNDSTRÖM, A. 2015. Covenants with broken swords: Corruption and law enforcement in governance of the commons. *Global Environmental Change,* 31**,** 253-262. 10.1016/j.gloenvcha.2015.02.002.
SUTER, J. F., VOSSLER, C. A. & POE, G. L. 2009. Ambient-based pollution mechanisms: A comparison of homogeneous and heterogeneous groups of emitters. *Ecological Economics,* 68**,** 1883-1892. 10.1016/j.ecolecon.2009.01.004.
SUTTER, M., HAIGNER, S. & KOCHER, M. G. 2010. Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies,* 77**,** 1540-1566. 10.1111/j.l467-937X.2010.00608.x.
SY, V. D., HEROLD, M., ACHARD, F., BEUCHLE, R., CLEVERS, J. G. P. W., LINDQUIST, E. & VERCHOT, L. 2015. Land use patterns and related carbon losses following deforestation in South America. *Environmental Research Letters,* 10. 10.1088/1748-9326/10/12/124004.
TACCONI, L., RODRIGUES, R. J. & MARYUDI, A. 2019. Law enforcement and deforestation: Lessons for Indonesia from Brazil. *Forest Policy and Economics,* 108. 10.1016/j.forpol.2019.05.029.
TAVONI, A., DANNENBERG, A., KALLIS, G. & LOSCHEL, A. 2011. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc. Natl. Acad. Sci. U.S.A.,* 108**,** 11825-9. 10.1073/pnas.1102493108.
THÖNI, C. 2013. Inequality aversion and antisocial punishment. *Theory and Decision,* 76**,** 529-545. 10.1007/s11238-013-9382-3.
TURUBANOVA, S., POTAPOV, P. V., TYUKAVINA, A. & HANSEN, M. C. 2018. Ongoing primary forest loss in Brazil, Democratic Republic of the Congo, and Indonesia. *Environmental Research Letters,* 13. 10.1088/1748-9326/aacd1c.
VANDER DONCKT, M., CHAN, P. & SILVESTRINI, A. 2021. A new global database on agriculture investment and capital stock. *Food Policy,* 100. 10.1016/j.foodpol.2020.101961.
VARUGHESE, G. & OSTROM, E. 2001. The Contested Role of Heterogeneity in Collective Action: Some Evidence from Community Forestry in Nepal. *World Development,* 29.
VELEZ, M. A., MURPHY, J. J. & STRANLUND, J. K. 2010. Centralized and Decentralized Management of Local Common Pool Resources in the Developing World: Experimental Evidence from Fishing Communities in Colombia. *Economic Inquiry,* 48**,** 254-265. 10.1111/j.1465-7295.2008.00125.x.

VOLLAN, B., PRÖPPER, M., LANDMANN, A. & BALAFOUTAS, L. 2019. Self-Governance and Punishment: An Experimental Study among Namibian Forest Users. *Economic Development and Cultural Change,* 67**,** 935-967. 10.1086/700098.
VOORS, M., TURLEY, T., KONTOLEON, A., BULTE, E. & LIST, J. A. 2012. Exploring whether behavior in context-free experiments is predictive of behavior in the field: Evidence from lab and field experiments in rural Sierra Leone. *Economics Letters,* 114**,** 308-311. 10.1016/j.econlet.2011.10.016.
VORLAUFER, M., IBANEZ, M., JUANDA, B. & WOLLNI, M. 2017. Conservation versus Equity: Can Payments for Environmental Services Achieve Both? *Land Economics,* 93**,** 667–688.
VOSTI, S. A., CARPENTIER, C. L., WITCOVER, J., VALENTIM, J. F. J. A. T. & DEFORESTATION, T. 2001. Intensified small-scale livestock systems in the Western Brazilian Amazon. 113-133.
WEBER, J. G., SILLS, E. O., BAUCH, S. & PATTANAYAK, S. K. 2011. Do ICDPs Work? An Empirical Evaluation of Forest-Based Microenterprises in the Brazilian Amazon. *Land Economics,* 87**,** 645–681.
WEST, T. A. P., BORNER, J., SILLS, E. O. & KONTOLEON, A. 2020. Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. *Proc Natl Acad Sci U S A,* 117**,** 24188-24194. 10.1073/pnas.2004334117.
WILSON, D. S., OSTROM, E. & COX, M. E. 2013. Generalizing the core design principles for the efficacy of groups. *Journal of Economic Behavior & Organization,* 90**,** S21-S32. 10.1016/j.jebo.2012.12.010.
WOLFERSBERGER, J., DELACOTE, P. & GARCIA, S. 2015. An empirical analysis of forest transition and land-use change in developing countries. *Ecological Economics,* 119**,** 241-251. 10.1016/j.ecolecon.2015.08.018.
WRIGHT, G. D., ANDERSSON, K. P., GIBSON, C. C. & EVANS, T. P. 2016. Decentralization can help reduce deforestation when user groups engage with local government. *Proc Natl Acad Sci U S A,* 113**,** 14958-14963. 10.1073/pnas.1610650114.
WU, J. & YU, J. 2017. Efficiency-Equity Tradeoffs in Targeting Payments for Ecosystem Services. *American Journal of Agricultural Economics,* 99**,** 894-913. 10.1093/ajae/aaw095.
WUNDER, S. 2015. Revisiting the concept of payments for environmental services. *Ecological Economics,* 117**,** 234-243. 10.1016/j.ecolecon.2014.08.016.
WUNDER, S., BROUWER, R., ENGEL, S., EZZINE-DE-BLAS, D., MURADIAN, R., PASCUAL, U. & PINTO, R. 2018. From principles to practice in paying for nature's services. *Nature Sustainability,* 1**,** 145-150. 10.1038/s41893-018-0036-x.
WUNDER, S., BÖRNER, J., EZZINE-DE-BLAS, D., FEDER, S. & PAGIOLA, S. 2020. Payments for Environmental Services: Past Performance and Pending Potentials. *Annual Review of Resource Economics,* 12**,** 209-234. 10.1146/annurev-resource-100518-094206.

Errata list:

Page	Line	Change from	Change to
31	7	“difference and difference”	“difference-in-difference”
37	26	“of the results. Levitt and List (2007) point (Levitt and List, 2007)out five important”	“of the results. Levitt and List (2007) point out five important”
168	11	“Farm income (ha)” “Environmental income (ha)” “Other income (ha)”	“Farm income (USD)” “Environmental income (USD)” “Other income (USD)”

Naime Sánchez-Henkel, Julia del Carmen

Julia was born in Mexico in 1991. She holds a bachelor degree in Biology from the National Autonomous University of Mexico (UNAM) and in Economics from New York University (NYU). She also holds a master degree in Environmental Economics and Climate Change from the London School of Economics and Political Science (LSE).

Her research is focused on the linkages between economic growth and deforestation, as well as on evaluating the effectiveness of conservation policies when implemented in different socioeconomic contexts. She has fieldwork experience in Mexico, Peru and Brazil.

The thesis consists of an introduction and four independent research papers. The two main objectives of the thesis are first, to evaluate the potential of collective Payments for Ecosystem Services (PES) to deliver on conservation and development outcomes under different contexts, and second, to improve our understanding of the causes of deforestation at the national level. The thesis draws from development, behavioural and experimental economics theories and methods.

Supervisors: Arild Angelsen and Frode Alfnes

E-mail: junaime@nmbu.no
Twitter: @julianasah

School of Economics and Business
Norwegian University of
Life Sciences (NMBU)
P.O Box 5003
N-1432 Ås, Norway

Telephone: +47 6496 5700
Telefax: +47 6496 5701
e-mail: *hh@nmbu.no*
http:/www.nmbu.no/hh

ISSN: 1894-6402
ISBN: 978-82-575-1881-3

ISBN: 978-82-575-1881-3
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

113096 / ANDVORDGRAFISK.NO