Simple incentives and intra-group dependence promote success in ecosystem services programs: Evidence from a framed field experiment in rural Lao PDR

Carl Salk

Maria-Claudia Lopez

Grace Wong

Introduction

When forest-derived greenhouse gas emissions are avoided, the benefits are felt worldwide, but the cost of reduced availability of forest resources is highly localized. Performance-based incentives, known as payments for ecosystem services (PES) are potential mechanisms to compensate local resource users for the costs they incur by changing their land use and forest management practices. In recent years, PES has emerged as a central mechanism to the 'reducing emissions from deforestation and forest degradation' (REDD+) concept to motivate forest management and conservation actions at local and national levels (Pattanayak et al. 2010; Engel et al. 2008). However, research into how best to deliver payments and whether schemes have a lasting post-incentive effect (either positive or negative) on the forest condition or user motivation is in its infancy (Wunder 2013). In this paper, we use a forest-framed field experiment to address these questions in the context of shifting cultivation landscapes in rural Lao PDR.

A key component of the design of any PES program is how to deliver benefits to local users. Direct cash payments are perhaps the simplest option, but payment in kind, insurance against crop failure, payments indexed to agricultural prices, and access to credit are other possibilities (Engel et al. 2015; Wong 2014; Chantharat, 2011; Pham 2013; To, 2012). Some schemes follow a combination of these options. A further dimension is whether to condition payments on individual-level behavior or to rely on a group-level metric, for instance the aggregate performance of a group of or a village (Alston et al., 2013; Loft et al., 2014). Empirical and experimental evidence about which of these options most effective is mixed (Narloch et al., 2012).

PES schemes are often criticized out of concern that they could replace resource users' intrinsic motivation to conserve resources with a mercenary attitude focused only on benefits derived from the scheme (Vatn, 2010; Muradian et al., 2013), a process commonly known as 'crowding out' (Frey and Jegen, 2001). Crowding out would have consequences for the long-term ability of PES to achieve policy goals. Once incentives end, for instance when funding or political will for a project runs out, targeted resource users may no longer have any reason to conserve, leading to poorer outcomes than before the program began. Indeed, some evidence shows this happening in experimental and real life settings (Cardenas et al., 2000; Frey and Jegen, 2001). However, it is equally plausible that external incentives would reinforce or even strengthen pre-existing motivations. This outcome, known as 'crowding in,' has also been observed in some experiments(e.g. Narloch et al, 2012; Lopez et al., 2013; Arriagada et al. 2015). A theory of when and where crowding out or crowding in is expected is not yet well articulated.

Careful tailoring of incentives is crucial when ecosystem-use practices to be incentivized are central to local people's livelihoods and cultural practices. In this study, we examine the impacts of different incentive types in the context of shifting cultivation, also known as 'swidden.' Swidden is practiced in most tropical regions worldwide (Barrera-Bassols and Toledo, 2005; Styger et al., 2007; CITATIONS) and is a vital source of food for millions of people (Minang et al., 2014). Its practitioners are frequently disadvantaged minority peoples with traditional land and resource rights that are recognized by themselves and their neighbors, but not always by the government Mertz et al. 2009; Padoch et al. 2007). However, misconceptions are common and swidden is often perceived as a driver of deforestation and a backward practice associated with low productivity and poverty, in spite of abundant evidence that management of dynamic forest and agriculture mosaics in a swidden landscape can maintain and enhance biodiversity and environmental services (Rerkasem et al. 2009, Ziegler et al. 2013, Magnuzsewski et al., 2015). In this context, PES incentives must be carefully targeted to reduce emissions while preserving livelihood provision and cultural practices.

In this study we implement a new framed field experiment in four swidden-practicing communities in northern Laos. We assess potential mechanisms for PES payments for their direct impacts and their potential for long-term behavior change (crowding in or crowding out). The game's dynamics reflect the reality that many rural people cannot completely eliminate their dependence on forests for shifting cultivation, but might reduce the intensity of their forest use if provided with assistance to manage environmental unpredictability. It also reflects the reality that incentive programs eventually end, and measures how players respond in a post-incentive period. We specifically test three different mechanisms to deliver PES: individually-directed payments (IP), group-based payments (GP) and insurance (INS). To our knowledge, no previous study has compared these PES mechanisms and their post-incentive efficacy in the context of shifting cultivation. Here, we seek to answer the following questions:

(1) Do individually-delivered payments, group-directed payments, or insurance better motivate changes in resource use patterns?
(2) Do the above PES mechanisms leave lasting changes in resource use patterns after incentive programs end?
(3) Do players' communication, motivation and perceived fairness of the games explain outcomes of the different mechanisms?
(4) How do individual-level attributes like age, leadership, resource dependence and education affect participation in PES schemes and their lasting effects after incentives end?

Our findings show that direct payments are more effective than insurance, and that group-level payments give better outcomes during the treatment rounds than individual payments. However, the lasting effect was minimal for all mechanisms, with neither crowding-out nor crowding-in observed. These results may be partially explained by increases in communication among the players and perceptions of fairness during the GP treatment; individual socio-economic traits had little apparent relation to within-game performance.

Methods

We address the questions outlined above with a framed field experiment in which groups of 8 participants made decisions about how much land to cultivate in a stylized forest sized to accommodate 100 potential agricultural plots. As in real-world shifting cultivation, the participants used a new area of forest in each round. In each round, each player chose how much to cultivate, with an upper limit of 10 patches per round. The players' earnings increased with the number of plots they cultivated, but decreased with the total number of plots used by the entire group. Earnings further depended on whether randomly generated rainfall was abundant or poor. An example of these calculations can be seen in Figure 1. The game was implemented in four rural communities in Phonxay District, Louangprabang Province in the Lao Peoples Democratic Republic. In each village, we conducted three games sessions with separate groups of 8 players In each session, we tested one of three PES mechanisms: a direct payment for moderating harvest levels known as the 'individual payment' (IP), a payment to each group member if the entire group moderated cultivation – the 'group payment' (GP,) and insurance against poor rainfall (INS). The game was divided into three stages: pre-incentive, incentive and post-incentive. The pre-incentive stage served as a baseline against which to measure two different effects, (1) the 'incentive effect', or the difference between pre- and during-incentive agricultural land use, and (2) the 'lasting effect', the difference between pre- and post-incentive land conversion (Figure 2). At the end of the experiment, all players were compensated with tangible goods proportioned to their in-game earnings.

Because the primary goal of this research was to assess the relative effectiveness of different PES interventions, the treatments were balanced so that they had identical harvesting levels to achieve the social optimum or Nash equilibrium, and identical expected payouts at the Nash equilibrium, and nearly identical expected payouts at the social optimum. The INS treatment had a social-optimal payout 4.3% higher than the GP and IP treatment. For practical reasons discussed in the supplement, perfectly identical expected payouts at the social optimum were not possible. This balance had the added benefit of ensuring equal earning opportunity among participants in the different treatments.

To provide additional information on interactions among the players, two people made independent observations according to a fixed protocol. These observations included several variables rated on an ordinal scale from 0 to 3, including how much players talked, both overall and specifically about topics like how much to harvest, how to maximize their payouts, criticism for overharvesting, or praising others' actions. The protocol also included an assessment of whether the group had a leader, and if so who that was. In addition to their data collection duties, the observers also helped players with their calculations and completion of score sheets as needed. Following the game, players were asked to respond to a short survey about their goals and decision-making process during the game. These included a question about whether the player was motivated by the incentive, and another about whether they perceived the incentive as fair. They were also asked how much time they spent in real-life forests. In addition to this information, we had demographic data about players (age, wealth, education, sources of livelihood, etc.) from a separate survey team that had visited the villages in the two weeks before the games took place.

Additional details of payouts under different scenarios, a fuller outline of the experimental game and analytical methods are found in the Supplemental Information.

Results

Pre-incentive (baseline) cultivation averaged 6.64 patches per round per participant (Figure 3A). This value is significantly higher than the social optimum of 6 (Mann-Whitney test, p=.0037), but still much closer to the cooperative optimum than to the Nash equilibrium of 10. As expected, during the pre-incentive stage there was no observed difference in cultivation among the incentives (Kruskal-Wallis test, p=.308). During the incentive stage, cultivation decreased for all PES schemes (Figures 3A, 3B). For the GP and IP incentives this effect was statistically strong (Mann-Whitney test, p<.0001 (GP) and p=.0004 (IP). However, for the insurance incentive, the effect was minimal (Mann-Whitney test, p=.326). These effects differed significantly from one another (Kruskal-Wallace test, p=.0005), with the GP and IP incentives showing a bigger effect than the insurance incentive (Figure 3C). While the PES incentive was in place, cultivation under the GP incentive was statistically indistinguishable from the cooperative optimum of 3 (Mann-Whitney test, p=.11), but the IP and INS incentives cultivation was significantly higher than this value (Mann-Whitney test, p<<.05). During the post-incentive stage, cultivation rates showed a rebound to levels similar to the pre-incentive stage; the lasting effect was minimal in all cases (Figures 3B, 3D). The GP incentive showed a trend toward lasting reductions in cultivation, but statistical evidence for this observation was equivocal (Mann-Whitney test, p=.1416). The IP and INS incentives' lasting effects were even closer to zero (Figure 3D; Mann-Whitney test, p=.387 (IP), p=.765 (INS)). The incentives' lasting effects did not differ significantly from one another (Kruskal-Wallis test, p=.255). Cultivation under all incentives was in the vicinity of the Pareto optimum of 6, however all groups were marginally different from this value (Mann-Whitney test, p=.109 (GP); p=.022 (IP); p=.071 (INS)). Although equitable use of the forest resource occurred sometimes, the harvesting decisions frequently deviated from the group mean decision for a given round across all treatments (Figure 3D).

Observations of the participants showed consistent decreases in communication among participants over the course of the experiment in the IP and INS incentives (Figure 4A). However, the GP incentive showed an uptick in communication during the incentive rounds (Figure 4A). Within experimental stages there was no discernable difference in communication among the incentives with the exception of the GP incentive during the incentive stage (permutation tests, p<.05). Participants in the GP incentive reported both significantly greater motivation (Figure 4B) and perceived fairness (Figure 4C) relative to the IP and INS incentives (exact binomial tests, p<.05).

Individual-level covariates (age, gender, years of education, land owned by one's family, the amount of time spent in the forest in the past month and observed in-game leadership) showed very little relationship with harvesting levels. When absolute pre-treatment harvesting decisions were used as the response variable, only time spent in the forest had a significant (and positive) predictor. However, when controls for villages were added to the model, this relationship fell away, apparently due to lower than average cultivation decisions in a villages whose inhabitants also spend less time than average in the forest. Cultivation decisions relative to other players in the same game showed no significant relationship with any of these variables (full results not shown).

Discussion

This research has demonstrated that a simple framed experiment can reveal differences in how residents of shifting cultivation communities change their behavior in response to different incentive structures. Overall, the participants from these four villages showed a relatively high level of cooperation. In no instance did any group fully defect to the Nash equilibrium strategy (cultivation of the maximum possible amount by all players). That group-level averages of cultivation were so close to the Pareto expectation in most stages and incentives (Figure 3B) is remarkable. However, these mean outcomes mask the substantial variation among players' decisions (Figure 3A), which show that free-riding did indeed take place in spite of seemingly cooperative group-level results. Our results show that resource user responses to incentives are not simply a function of potential financial gain. The payoff structures of the three incentive schemes were quite similar, particularly in the case that players were cooperative. However, our results showed that a lump-sum payment that depended on group-level choices (GP) was more effective than an otherwise identical payment that required only an individual-level reduction in cultivation (IP). In turn, both of these direct payment incentives were better than an insurance incentive (INS) which yielded no discernable change in cultivation practices.

That a group-level payment was more effective than an individual-level payment is congruent with some, but not all, related research. Intra-group dependency is a factor promoting successful common pool resource management outcomes (Frey and Rusch, 2013). By introducing a dependency among the players, a group payment increases the loss inflicted on others when a player defects from the cooperative strategy without increasing the benefit to the free-rider. It is possible that this dynamic is behind the observed increase in discussion observed when the GP incentive is in effect and the increased motivation reported in post-experiment surveys. It is less obvious how this dynamic would lead to the greater perceived fairness of the GP treatment except through indirect pathways, for instance if increased communication led to better outcomes which the players in turn perceived as fair. Alternatively, the GP treatment may have succeeded simply by signaling to players that this is a group activity and that they should work together, an interpretation that would also be consistent with the increased communication seen during the GP rounds. Regardless of the particular explanation, it is important to note that at least one experimental study has found the opposite of this result. Narloch et al. (2012), using an agriculture-framed common pool resource game in the highlands of Peru and Bolivia, found that individual rewards promoted conservation action more effectively than group rewards. The immediate reason for these contrasting results is not clear. Many different factors have been shown to affect the cooperation of users in both real-world (Ostrom, XXXX) and game-based (CITATION) resource allocation dilemmas, making it difficult to speculate about the cause of these contrasting outcomes.

Why the insurance incentive was ineffective is not immediately clear. In theory, it provides nearly identical expected benefits to the direct payout incentives. Part of the problem may be the relative unfamiliarity of insurance to rural Lao residents, a situation that the Lao PDR government aims to change (Lao PDR Ministry of Agriculture, 2010). Further, the relative complexity of this incentive, combined with the low numeracy of many players, could mean that the benefits of this instrument were not apparent to many of the participants. Indeed, some case studies suggest that complexity of institutions and incentives can lead to confusion and indifference on the part of participants (e.g. Scheberle, 2000). This result does not mean that insurance cannot be an effective incentive for shifting cultivators, but it does suggest that additional effort is necessary to help participants understand insurance programs. Games like ours may be helpful training tools for this purpose

In spite of the significant incentive effect for the direct payout groups, it is important to note that no treatment showed an unequivocal lasting effect once the incentive program ended. That the group payment incentive came the closest to providing a lasting impact is tantalizing, but ultimately an inconclusive result, illustrating the difficulty of achieving lasting impacts from projects that have a finite time-frame and budget. However, this result should not be viewed in an entirely negative light. While no treatment had a clear lasting reduction in cultivation, neither did any treatment lead to a lasting *increase* in cultivation. We saw absolutely no evidence of crowding out in this study.

Reduced dependence on forest for shifting cultivation could result in more forests and fallows within the landscape via two non-exclusive routes. One is less land dedicated to shifting cultivation, and hence more land reserved for forest or other uses (land sparing). The second is less frequent rotation, and hence an older average age of fallows. Both mechanisms would bring benefits for carbon sequestration, biodiversity and other ecosystem services. While our game does not address these possibilities, both are consistent with our results.

As in all game based studies, the question of real-world validity of results is difficult to definitively address. That players who visit real forests less also harvest less in the context of the game is suggestive of behavioral validity of the experiment (sensu Handberg and Angelsen, 2015), although due to the covariance of these variables among villages this finding is not completely statistically conclusive. These findings suggest some application to real-world settings, not just where shifting cultivation is practiced, but anywhere rural resource users are the targets of incentive programs. Certain features may be predictive of success, particularly simplicity of incentives and intra-group dependence. The details of how the reductions in harvesting seen in the game would play out in practical applications are beyond the scope of this study. Regardless, these findings show that swidden communities are fully capable to participate in incentive programs designed for more sustainable forest and land use management through collective decision-making. The potential adaptation of the incentives within the dynamism of their actual land use practice remains to be seen and warrants further social-ecological research to understand how these incentives can sustain provision of ecosystem services from these landscapes over the long-term.

Detailed Methods

The framed field experiment

We designed an appropriation experiment centered around a shared group resource (framed as a forest used for shifting cultivation) in which players get private benefits from extracted resources, but as a group they gain from the units of the resource left untouched. However, unlike previous linear appropriation games (Travers et al., 2011; Blanco et al., 2015; Lopez et al., in prep) players were not paid directly for unharvested resources. Rather, the value of extracted resources (cultivated plots) was proportional to the amount of un-extracted resource (intact forest). This dynamic reflects some real-world characteristics of shifting cultivation; no harvest is possible if no land is planted, but over-use of land can reduce the yield of cultivated areas. The game consists of 24 rounds (each analogous to an agricultural season), separated into three eight-round stages. A new forest area is used for cultivation in each round, so there are no historical effects of decisions in previous rounds.

Each group is composed of eight members of the study village who were selected with the aim of including those who practice shifting cultivation and balancing male and female and otherwise representing the diversity (age, ethnicity, etc.) of the village. However, practical considerations did not always allow groups that were as well-mixed as we would have desired. The group shares a common forest containing 100 patches that can potentially be cleared and farmed. Open communication among the players is allowed throughout the game. Participants are told at the beginning that there are no neighbors who may encroach or steal from their land, and no government or other authority to impose restrictions on them. Each participant must decide how many patches of forest to clear and plant with crops, with a lower limit of 0 and an upper limit of 10. Each player's total income is the product of three factors: how many patches he/she cultivates, the number of patches left untouched by the group as a whole, and randomly generated weather. Figure 1 in the main text shows an example harvest calculation.

To compute the number of un-cultivated patches, the eight players' individual cultivation decisions are summed and subtracted from 100 (the total number of patches in the forest). The maximum group-level cultivation is 80 patches (10 patches x 8 players), ensuring that a minimum of 20 patches are left uncultivated in each round. This also ensures players a minimum per-patch harvest from their cultivation. The dependence of per-patch agricultural productivity on un-cultivated forest is explained in terms of ecosystem services; the un-cut forest provides services such as watershed protection, landslide prevention and pollination that increase the productivity of farmed patches. This mechanism sets up a tradeoff between individual cultivation (more patches cultivated mean increased harvest) and group cultivation (more patches cut means reduced ecosystem services and thus decreased harvest). The final factor affecting cultivation yield is unpredictable weather. This factor mimics the real-world concerns of shifting cultivators, and adds an element of chance needed for the insurance incentive (described below). Weather was announced at the end of each round. In the case of poor rainfall, players' scores for the round remained unchanged. With good rains, the scores were multiplied by three.

The game is structured in three stages to measure baseline behavior, the direct impact of incentives, and the lasting effect of incentives (either positive or negative) after they are removed. In the first eight rounds (stage 1) the game is played exactly as described in the previous paragraphs; there is no institution in use to manage the CPR other than what the players arrange among themselves. At the beginning of stage 2 (rounds 9-16), the moderator explains that an outside organization interested in forest conservation has made an offer in exchange for certain harvesting patterns. This 'offer' is one of the three incentives described below. In the third and final stage the incentive ends and the game proceeds as in the first stage.

During stage 2 of the game, one of three possible incentives is implemented. These are designed to assess the responses of shifting cultivation-dependent people to PES interventions. The incentives are a payment offered for individual limitation of cultivation (individual payment – IP), a payment dependent on group-level cultivation decisions (group payment – GP) and insurance against weather-related crop failures (insurance – INS).

Under the IP incentive the organization pays each player 200 points per round if he or she cultivates no more than three patches during that round. This bonus is paid regardless of the decisions of the other participants or the weather. The GP incentive works the same as the IP, except that the payment is only made if the entire group cultivates no more than 24 patches for agriculture. If the 24 patches are equitably distributed, then each player can cultivate three patches and still receive the bonus. However, whether the payout is made does not depend on equity; the organization will pay 200 points to each participant regardless of their individual decisions as long as the group total cultivation does not exceed 24 patches. The INS incentive offers players a buffer against unpredictable weather. This incentive functions at the individual level. Each player who cultivates no more than three patches receives an effective guarantee of good weather. If the randomly-generated weather is good, then the player receives only the standard payout from cultivating and harvesting crops. However, if the weather is poor, the player receives a payout equivalent to the difference between their actual harvest and the amount they would have harvested had the weather been good. This effectively means their total payout for the round is tripled. Players who cultivate more than three patches receive whatever harvest results from their planting, but do not receive any insurance payout in the event of poor weather.

The game is designed so that the payouts at the cooperative social optimum are virtually identical across incentives. For the control rounds (stages 1 and 3), the Pareto optimum was to cultivate six patches per participant. This gives an expected payout of 624 points per player-round (312 points under bad weather, 936 points with good weather; the two weather outcomes are equally likely). Thus, under perfectly cooperative play, a player could expect to earn a total of 9,984 points during stages 1 and 3. Optimal payouts were designed to be similar among the three types of incentive rounds. Under the GP incentive, cultivation of three patches per person led to the highest group-total earnings, of 656 points per round (428 under bad weather, 884 with good weather). The IP incentive also had a near-optimum with the same earnings as the GP incentive when three patches were cultivated per person; no higher equally-shared earnings are possible. However, IP incentive had a cryptic optimum with slightly higher total earnings, but with substantial inequality among players. This optimum consisted of three players cultivating the maximum of 10 plots and all others harvesting 3. Under this scenario the players harvesting 10 would earn an expected 1100 points, and those harvesting three would earn 530 points, for a group-level mean of 743.75 points/player. However, in the field, the players never discovered this optimum. Under the insurance incentive, the social optimum was cultivating three patches per person. This led to an expected payout of 684 points per round which was the same regardless of good or bad weather. Under the control and all three incentives the Nash equilibrium was cultivation of 10 patches per player. In all cases this led to expected earnings of 400 points per round (200 under bad weather, 600 under good weather).

Free riding was also a possibility in this game. If all players cooperated but one, the payout under average weather in control rounds for a free-rider cultivating 10 patches was 960 points/round, with the income of cooperating players (who still cultivated 6 plots) reduced to 576 points. In the incentive rounds the potential benefit of free-riding was even greater. Under the INS incentive, a free-rider cultivating 10 patches would receive 1380 points, while the seven cooperating players cultivating 3 patches would receive 621 points. With the IP incentive, the free-rider would also receive 1380 points and the cooperators 614 points per round. This gap was larger in the GP incentive due to the loss of the bonus payment by all players if anyone defected; the free-rider still earns 1380 points, but the cooperators earn only 414 points under average weather.

Game implementation

The game was implemented in four villages (described above) during January 2015. In each village, the game was played with three distinct groups of eight people. The field team worked with village leaders to ensure that close relatives were not included within any of the groups. The game was held in a public place, often a school or a community hall. The team that implemented the game was the same in all communities and was composed of one moderator (female) and two observers (both male) who also helped less literate participants fill in their score sheets. The game team was trained in English, but they then practiced running the game in Lao before the field implementations. All field implementations were in the Lao language. Although some players were native speakers of other languages, all understood and spoke Lao well. During the activity, players were seated in a circle, spaced so they could easily speak to one another, but so they could not easily see what is written on one-another's score sheets. If players attempted to look at others' score sheets or display their own, they were gently but firmly told not to do so. The game typically took 1.5-2.5 hours to play, not counting the post-game survey; the slower groups were generally those who required more assistance with score sheets.

Instructions were communicated aloud to all participants in each stage of the game. The first two stages were preceded by three practice rounds each which give participants practical examples so they better understood the mechanism of the game. During the three stage 1 practice rounds, the players were instructed to cultivate 3, 7 and 10 patches respectively. In the three stage 2 practice rounds, the moderator told the players to cultivate 2, 3 and 4 patches respectively. In case of questions, the question and answer were explained to all participants. Participants' decisions were made in secret by marking the number of patches cultivated on a slip of paper. This slip represents all possible decisions (from 0 to 10 patches harvested) both in numbers and visually to help players who may not be familiar with written numbers. The moderator summed up the individuals' decisions and announced them publicly. In each practice round, the moderator told the players how many patches of forest to clear and cultivate, and checked to see that they fill out their decision slips and score sheets correctly. The harvesting levels in the practice rounds are chosen to bracket the group-level harvesting optimum and help the players to discover this without directly giving it away.

Each stage used a separate score sheet to reduce confusion during introduction and removal of incentives. Players fill out the score sheets themselves, but the moderator and observers check to ensure that they are filled out correctly and honestly. In the field no attempts at cheating were seen, but many participants with low education levels, particularly the elderly, needed assistance tallying their scores. Occasional mistakes were discovered during data entry, however these mistakes had minimal impact on the overall results.

Randomly-generated weather was announced at the end of each round, after cultivation decisions were submitted and the group-level harvest announced. To generate the weather outcome, the moderator placed her hands behind her back with a piece of candy in each of her hands. One candy had a blue wrapper, representing abundant rain, and the other a brown wrapper, representing failure of rain and a poor harvest. The moderator offered her two closed hands to a participant who chose one. If the chosen hand contained the green candy, all participants multiplied their base harvest (the product of patches cultivated individually and patches left un-cultivated by the group) by three. In the next round, the next player to the right made the selection, working around the circle so that all players participated equally.

To give the players an incentive to take the game seriously and try to optimize their decisions players were given rewards proportional to their points earned. During the initial instructions it was explained to the players that the points earned during the 24 game rounds (practice rounds were excluded from earnings) would be redeemed for rewards at the end of the game. The rewards were 120 gram packets of laundry washing powder. Before the game it was explained that every 2000 points earned in the game was equivalent to one packet. To emphasize the tangible rewards, a case of the packets was left visible on the floor during the game. These values were always rounded up to the next full packet; for instance, a player earning 8,463 points would receive 5 packets at the end of the game. In the villages where these games took place this type of washing powder is used by all households and both male and female participants appeared to appreciate them. The bags had a local retail price of 3000 Lao kip each, equivalent to about $US 0.38 at the time of field implementation. Thus, the expected payout for the individual and group bonus incentives under perfect cooperation and equity (and average weather) was 624 points/round x 16 rounds + 656 points/round x 8 rounds = 15,232 points, or eight bags of laundry powder. Under the INS incentive the social optimum payout given average weather was 624 points/round x 16 rounds + 684 points/round x 8 rounds = 15,456 points, also equivalent to eight bags of laundry powder. The Nash equilibrium for all incentives was 400 points/round x 24 rounds = 9,600 points, or five bags of washing powder.

Statistical methods

The two key response variables in this study are what we call the 'incentive effect' and the 'lasting effect' (Figure 2). The incentive effect is the extent to which cultivation changes between the pre-incentive (stage 1) and incentive (stage 2) rounds. The lasting effect is the difference in cultivation between the pre-incentive and post-incentive (stage 3) rounds. Unless stated otherwise, these values were calculated at the player level by averaging each player's decisions across the 8 rounds of each stage, and then taking the appropriate differences. All subsequent statistical analyses were conducted with replication of n=32 (8 players x 4 villages) per incentive. All tests described below were implemented in R version 3.0.3 (R Core Team, 2014).

Financial and time constraints allowed implementation of this game in four villages. For this reason, the unit of replication in this study was the player. That the players were nested within villages violates the independence assumptions of many statistical tests. Therefore, we tested for significant village-level effects using generalized linear models with dummy variables for treatments and villages. In all cases, estimates of village-specific coefficients were not significant at $p < .05$ (Table S2) and residuals were normally distributed (Shapiro-Wilks test, $p >> .05$). Thus, we proceeded with our analyses using standard non-parametric tests with the player as the unit of replication.

Normality of the cultivation distributions was evaluated using the Shapiro-Wilk test. Cultivation rates within each of the three stages and within all but one of the nine stage-incentive combinations were distinctly non-normal ($p<.05$), so non-parametric tests were used in all subsequent analyses. Under the assumption that groups were cooperating to maximize their earnings, cultivation was predicted to either be 6 for non-incentive rounds and 3 for incentive rounds (the Pareto optimum). Differences from the expected Pareto values were assessed using one-sample, two-tailed Mann-Whitney U-tests (a non-parametric analogue to the t-test). Alternatively, cultivation of 10 patches would be consistent with the Nash equilibrium expectation of economic theory. However, because this is the highest possible harvesting value, distributions in this range would be truncated, so it is not possible to use conventional statistics to test whether cultivation differs significantly from this value. However, we pursue other lines of reasoning to address this question. Differences among incentives in cultivation, incentive effect and lasting effect were assessed using the Kruskal-Wallace test, a non-parametric alternative to ANOVA.

Because the observations of among-player communication are ordinal (on a scale of 0, 1, 2 or 3) we used a permutation test to check for differences among the incentives. Tests were performed pairwise between each pair of treatments within each game stage for a total of nine tests. Comparisons among stage-treatment combinations were not performed because there were 36 possible combinations. For each permutation test, all communication values belonging to the two treatments being compared were randomized into two groups of 32 each. The difference between the two groups' means was computed and stored. This process was repeated 10,000 times to simulate a distribution of expected differences between the two groups under the assumption that the two treatments values come from the same underlying distribution. The quantile of the observed communication mean was calculated as a measure of how likely the observed communication differential was to be a result of chance. Since this is a two-tailed test, quantiles of <.025 or >.975 were considered to be significant at the $p<.05$ level.

Differences among the proportion of players in different incentives who reported that they were motivated by the incentive and who felt that the incentive was fair were assessed using exact binomial tests.

Acknowledgements

This work has been funded by UK aid from the UK Government (contract # …) and an NSF Coupled Natural-Human Systems grant (DEB-1114984) to Krister Andersson. The views expressed do not necessarily reflect the UK Government's official policies. We would like to thank Glenn Wright for statistical advice and Arlid Angelsen for helpful feedback on a draft of this manuscript.

References

Alston, L. J., Andersson, K., & Smith, S. M. (2013). Payment for Environmental Services: Hypotheses and Evidence. Annual Review of Resource Economics, 5(1), 139–159.

Anderies, J.M., Janssen, M.A., Bousquet, F., Cardenas, J.C., Castillo, D., Lopez, M.-C., Tobias, R., Vollan, B.R., and Wutich, A., 2011. The challenge of understanding decisions in experimental studies of common pool resource governance. Ecol Econ, 70(9), 1571-1579.

Arriagada RA, Sills EO, Ferraro PJ, Pattanayak SK (2015) Do Payments Pay Off? Evidence from Participation in Costa Rica's PES Program. PLoS ONE 10(7): e0131544.

Aswani, S., Gurney, G. G., Mulville, S., Matera, J., & Gurven, M. (2013). Insights from experimental economics on local cooperation in a small-scale fishery management system. Global Environmental Change, 23(6), 1402-1409

Barrera-Bassols N., Toledo V.M. 2005. Ethnoecology of the Yucatec Maya: Symbolism, knowledge and management of natural resources. J. Lat. Am. Geogr. 4:9–41.

Bruun TB, Egay K, Mertz O and Magid J. 2013. Improved sampling methods document decline in soil organic carbon stocks and concentrations of permanganate oxidizable carbon after transition from swidden to oil palm cultivation. Agriculture, Ecosystems & Environment 178:127–34.

Balliet, D. (2010). Communication and cooperation in social dilemmas: a meta-analytic review. Journal of Conflict Resolution, 54(1), 39-57. doi: 10.1177/0022002709352443

Bruun, T., A. de Neergaard, D. Lawrence and A.D. Ziegler (2009) Environmental consequences of the demise in swidden cultivation in Southeast Asia: Carbon storage and soil quality, Human Ecology 37: 375–388.

Cardenas, J. C., Stranlund, J., & Willis, C. (2000). Local Environmental Control and Institutional Crowding-Out. World Development, 28(10), 1719–1733.

Chantarat, S., Barrett, C. B., Janvilisri, T., Mudsri, S., Niratisayakul, C., & Poonswad, P. (2011). Index insurance for pro-poor conservation of hornbills in Thailand. Proceedings of the National Academy of Sciences, 108(34), 13951-13956.

Cinner, J.E., Folke, C., Daw, T., and Hicks, C.C., 2011. Responding to change: using scenarios to understand how socioeconomic factors may influence amplifying or dampening exploitation feedbacks among Tanzanian fishers. Global Environ Chang, 21(1), 7-12.

Dercon, S. (2002) Income Risk, Coping Strategies and Safety Nets. WIDER Discussion Papers, World Institute for Development Economics Research (UNU-WIDER), No. 2002/22, ISBN 9291901636.

Engel, S., Pagiola, S. & Wunder, S. (2008). Designing payments for environmental services in theory and practice: an overview of the issues. Ecol. Econ., 65(4), 663–675.

Engel, S., Palmer, C., Taschini, L., & Urech, S. (2015). Conservation Payments under Uncertainty. Land Economics, 91(1), 36-56.

Fox J, Fujita Y, Ngidang D, Peluso N, Potter L, Sturgeon J and Thomas D. 2009. Policies, political-economy, and swidden in Southeast Asia. Human Ecology 37:305–22.

Frey, BS and R Jegen. 2001. Motivation crowding theory. Journal of Economic Surveys 15:589-611.

Frey, U. J., and H. Rusch. 2013. Using artificial neural networks for the analysis of social-ecological systems. Ecology and Society 18(2): 40.

Gelcich, S., R. Guzman, C. Rodriguez-Sickert, J. C. Castilla, and J. C. Cárdenas. 2013. Exploring external validity of common pool resource experiments: insights from artisanal benthic fisheries in Chile. Ecology and Society 18(3): 2. http://dx.doi.org/10.5751/ES-05598-180302

Gurven, M., and J. Winking. 2008. Collective action in action: prosocial behavior in and out of the laboratory. American Anthropologist 110:179-190. http://dx.doi.org/10.1111/j.1548-1433.2008.00024.x

Handberg, Ø.N., Angelsen, A. (2015). Experimental tests of tropical forest conservation measures. Journal of Economic Behavior and Organization, in press.

Hett, C., Castella, J.C., Heinimann, A., Messerli, P., & Pfund, J.L. (2011). A landscape mosaics approach for characterizing swidden systems from a REDD+ perspective. Applied Geography, 32, 608–618.

Kurzban, R., Houser, D., 2005. Experiments investigating cooperative types in humans: a complement to evolutionary theory and simulations. Proceedings of the National Academy of Sciences 102, 1803–1807.

Lao PDR Ministry of Agriculture and Forestry (2010). Strategy for Agricultural Development 2011 to 2020 (Lao PDR), 67 pages.

Loft, L., Thuy, P. T., & Luttrell, C. (2014). Lessons from Payments for Ecosystem Services for REDD+ Benefit-Sharing Mechanisms. Bogor, Indonesia: Center for International Forestry Research (CIFOR).

Lopez, M. C., Murphy, J. J., Spraggon, J. M., & Stranlund, J. (2013). Does Government Regulation Complement Existing Community Efforts to Support Cooperation? Evidence from Field Experiments in Colombia. In J. List & M. Price (Eds.), Handbook on Experimental Economics (pp. 346–366). the Environment. Edward Elgar Publishing.

Magnuszewski P., Ostasiewicz K., Chazdon R., Salk C., Pajak M., Sendzimir J. Andersson K., 2015. Resilience and Alternative Stable States of Tropical Forest Landscapes under Shifting Cultivation Regimes. PLoS ONE 10(9): e0137497.

Mertz, O. Trends in shifting cultivation and the REDD mechanism (2009). Current Opinion in Environmental Sustainability, 1:156–160.

Minang P.A., van Noordwijk M., Kahurani E., editors. (2014). Partnership in the Tropical Forest Margins: a 20-Year Journey in Search of Alternatives to Slash-and-Burn. Nairobi, Kenya: ICRAF. pp 241.

Morduch, J. (1995). Income smoothing and consumption smoothing. Journal of Economic Perspectives, 9:103–114.

Muradian, R et al. (2013) Payments for ecosystem services and the fatal attraction of win-win solutions. Conservation Letters 6:274-279

Ostrom, E. (1990). Governing the commons: The evolution of institutions for collective action. New York: Cambridge University Press.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: self governance is possible. The American Political Science Review, 86(2), 404-417.

Ostrom, E., Gardner, R., & Walker, J. M. (1994). Rules, games, and common-pool resources. Ann Arbor: University of Michigan Press

Pattanayak SK, Wunder S and Ferraro PJ. 2010. Show me the money: Do payments supply environmental services in developing countries? Review of Environmental Economics 4(2):1–21.

Padoch, C., Coffey, K., Mertz, O., Leisz, S., Fox, J., and Wadley, R. L. 2007. The demise of swidden in Southeast Asia? Local realities and regional ambiguities. Geografisk Tidsskrift-Danish Journa lof Geography 107: 29–41

Paxson, C.H. (2000). Using weather variability to estimate the response of savings to transitory income in Thailand. In P. Bardhan and C. Udry (Eds), Readings in Development Economics, Volume 2: Empirical Microeconomics, Cambridge, Massachusetts, MIT Press.

Pham TT, Bennett K, Phuong VT, Brunner J, Dung LN and Tien ND. 2013. Payments for Forest Environmental Services in Vietnam: From Policy to Practice. Brief No. 22. Bogor, Indonesia: CIFOR.

Poteete AM, Janssen MA, Ostrom E: Working Together: Collective Action, the Commons and Multiple Methods in Practice. Princeton University Press; 2010.

Pravongviengkham, P.P. 1998. The Role of Animal Husbandry and Aquaculture in Improvements of Swidden-Based Livelihood Systems in the Lao PDR. PhD Dissertation. Bangkok, Thailand: Asia Institute of Technology.

R Core Team 2014 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rustagi, D., Engel, S., Kosfeld, M., 2010. Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management. Science 330, 961–965.

Rerkasem, K., D. Lawrence, C. Padoch, D. Schmidt-Vogt, A. D. Ziegler, and T .B. Bruun (2009). Consequences of swidden transitions for crop and fallow biodiversity in Southeast Asia. Human Ecology 37(3): 347–360.

Sally, D. (1995). Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. Rationality and Society, 7(1), 58-92. doi: 10.1177/1043463195007001004

Scheberla, D (2000). Moving toward community-based environmental management. American Behavioral Scientist, 44: 565-579.

Styger E., Rakotondramasy H.M., Pfeffer M.J., Fernandes E.C.M., Bates D.M. 2007. Influence of slash-and-burn farming practices on fallow succession and land degradation in the rainforest region of Madagascar. Agric. Ecosyst. Environ 119:257–269.

To P.X., Dressler W.H., Mahanty S., Pham T.T. and Zingerli C. 2012. The prospects for payment for ecosystem services (PES) in Vietnam: A look at three payment schemes. Human Ecology 40: 237–49.

Townsend, R. M. (1995). Consumption insurance: An evaluation of risk-bearing systems in low-income economies. Journal of Economic Perspectives, 9: 83–102.

Travers, H., Clements, T., Keane, A., Milner-Gulland, E. (2011) Incentives for Cooperation: The Effects of Institutional Controls on Common Pool Resource Extraction in Cambodia. Ecological Economics 71: 151–161.

Vatn, A. (2010). An institutional analysis of payments for environmental services. Ecological Economics, 69(6), 1245–1252.

Wong, G. (2014). The experience of conditional cash transfers: Lessons for REDD+ benefit sharing. CIFOR Infobrief no. 97. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Wunder, S. (2008). Payments for environmental services and the poor: concepts and preliminary evidence. Environ.Devel. Econ., 13(3), 279–297.

Figures

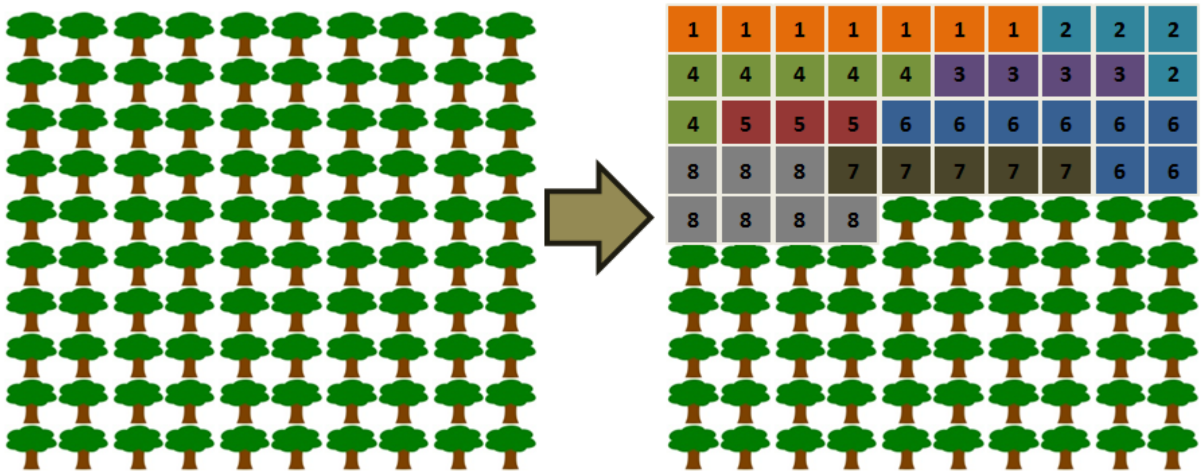

Figure 1. An example round of the shifting cultivation game. The left side of the figure represents the forest before cultivation with 100 available patches. The right side shows a hypothetical round of cultivation with a different color for the patches belonging to each player. The total cultivation by all players in this example is 44 patches, meaning 56 patches remain in the forest. In this example player 1 would earn 7 x 56 x 1 = 392 points if rainfall is poor, or 7 x 56 x 3 = 1176 if the rainfall is abundant. Full details of the game mechanics and payoff structure are found in the Supplemental Information.

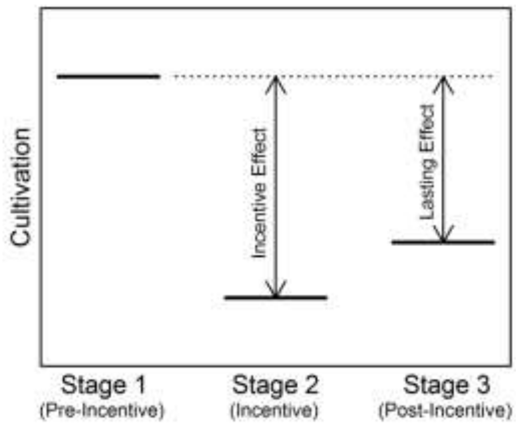

Figure 2. An idealized diagram showing the basic structure and outcome variables in the shifting cultivation game. The game is divided into three stages of eight rounds each. In stages 1 and 3 there is no external incentive to limit cultivation, although a cooperative group seeking to maximize their earnings would still do so. During stage 2 one of three possible incentives is offered as a means to reduce cultivation. Cultivation levels in stage 1 are the baseline against which choices in stages 2 and 3 are measured. This baseline allows direct comparison of groups that have a different inherent cultivation levels. We use the term 'incentive effect' to refer to the impact of the incentive while it is in effect (stage 2), and 'lasting effect' to mean the impact of the incentive after it has ended. While this diagram shows both effects as reducing cultivation (as hypothesized), it is possible for either or both effects to be in the opposite direction.

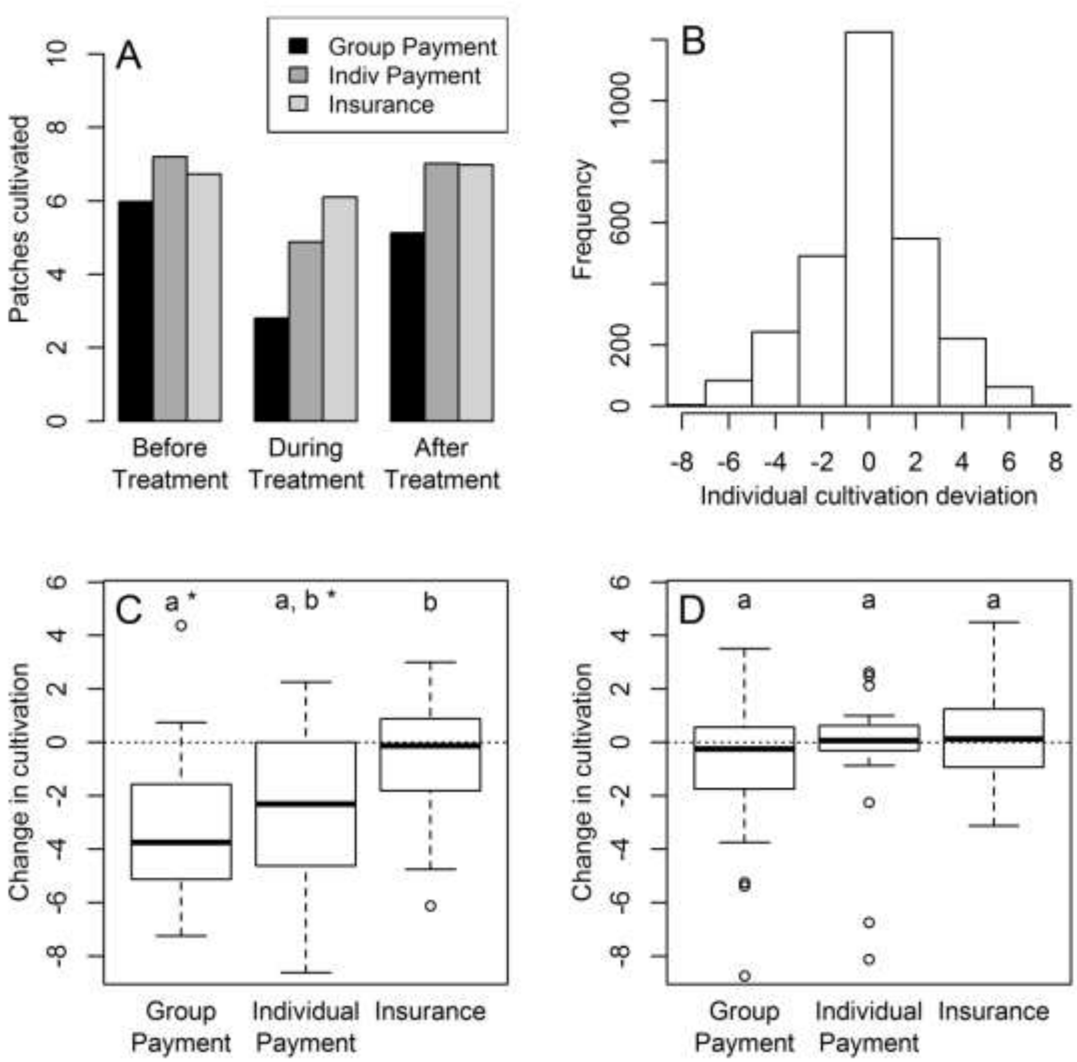

Figure 3. (A) The frequency of cultivation anomalies (the difference between an individual player's cultivation and the mean cultivation level for an entire group in that round) in all villages, incentives and rounds of the shifting cultivation game. (B) Mean cultivation as a function of game stage and incentive type. The incentive is only applied during stage 2. (C) The incentive effect (change in cultivation) between the first (pre-incentive) and second (incentive) game stages. Negative numbers mean harvesting decreased relative to the pre-incentive rounds. The letters at the top of the figure show statistical differences at the $p<.05$ level assessed using a Mann-Whitney test. Stars indicate groups whose mean was significantly different from zero (Mann-Whitney test with $p<.05$). (D) The lasting effect of after incentives end. The letters at the top of the figure show statistical differences at the $p<.05$ level assessed using Tukey's honestly significant difference test. No groups had a lasting effect that was statistically different from zero at the $p=.05$ level. The group payment incentive was closest with $p=.142$.

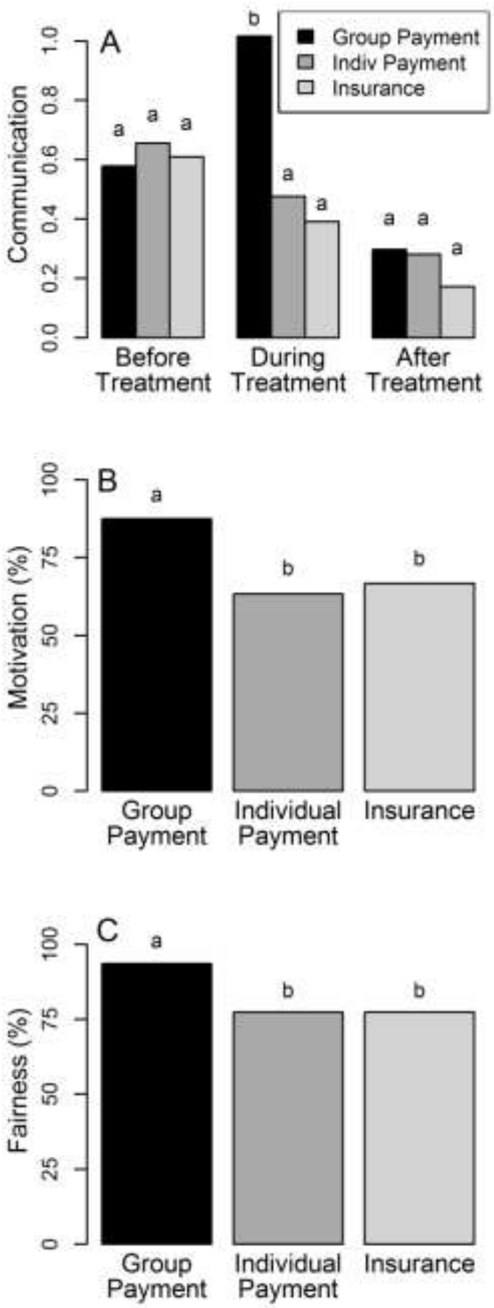

Figure 4. (A) Communication among players as a function of incentive type and experimental stage. Communication was assessed by two independent observers. Small letters refer to significant differences (see supplement) among incentives within each stage (before, during and after incentive), but *not* across stages. (B) Percentage of players reporting in a post-game survey that the incentive scheme motivated them to cultivate fewer patches. (C) Percentage of players reporting in a post-game survey that they felt the incentive scheme was fair. In B and C, small letters indicate significant differences ($p<.05$) among groups as determined using exact binomial tests (see supplement).