AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Ex ante evaluation of policy measures: Testing effects of reward or punishment with different probabilities in a framed field experiment about fertilizer reduction in palm oil production

Stefan Moser and Oliver Mußhoff[1]

Abstract

Palm oil production creates negative externalities, e.g., through intensive fertiliser application. Policies to limit externalities need an effective, sustainable and efficient measure We use a business simulation game in a framed field experiment in Indonesia to test ex ante different incentives for reducing such negative externalities. This setting allows inclusion of adequate contextual features, required for reasonable ex ante evaluation of policy measures. The different designs of the test incentives (either a reward or punishment) varied in their magnitude and probability of occurrence but with constant effects on expected income. Results show that participants react differently to these incentives, indicating that the design can contribute significantly to effectiveness, sustainability or efficiency. A high reward with a low probability was found to be the most effective and sustainable incentive. Moreover, for the most efficient design, a low and certain reward is indicated.

Keywords

Ex ante policy impact analysis, reward, punishment, framed field experiment, business simulation game, palm oil production, Indonesia, fertiliser use

JEL classifications: C91, Q18, Q52

1. Introduction

The use of palm oil is widespread. It can be found, for example, in food products, cosmetics or as biofuel. With a production volume of approximately 50 million tons in 2012, it is the most significant vegetable oil in the world (FAOSTAT, 2013). Starting in the 1960s, with a worldwide production of several million tons, palm oil production has grown exponentially, almost doubling every 10 years. With 23.6 million tons in 2012, Indonesia is the largest palm

[1] Stefan Moser and Oliver Mußhoff are with the Department of Agricultural Economics and Rural Development, University of Göttingen, Germany. E-mail: smoser@gwdg.de for correspondence. This study was financed by the Deutsche Forschungsgemeinschaft (DFG) within the collaborative German-Indonesian research project CRC990. The authors thank their Indonesian as well as their German partners for their generous support. Moreover, we thank the Editor and anonymous reviewers for their helpful comments and suggestions.

oil producer in the world and is working towards increasing its production by as much as 40 million tons in the year 2020 (UNCTAD, 2013).

However, there are negative externalities caused by palm oil production. Koh and Wilcove (2008) state that the expansion of palm oil production leads to deforestation, as well as to significant losses in biodiversity. Fitzherbert et al. (2008) additionally describe effects on habitat fragmentation and pollution. Reijnders and Huijbregts (2008) found significant CO_2 emissions caused by palm oil production, along with other factors caused by intensively applying fertiliser. The exaggerated fertiliser use leads to further negative externalities. For example, Sekhon (1995) shows that developing countries with an increased usage of N-fertiliser have more problems with groundwater pollution. Furthermore, the use of fertiliser in humid, tropical climates can increase NO_x emissions, which are as major contributors to global warming (Keller and Matson, 1994; Veldkamp and Keller, 1997; Palm et al., 2002; Veldkamp et al., 2008). Moreover, fertiliser can cause ground-level ozone in tropical oil palm plantations, high concentrations of which can be detrimental to human health (Hewitt et al., 2009). To limit these externalities from the extensive use of fertiliser in palm oil production, an effective, sustainable and efficient incentive system seems desirable. The term ‘effectiveness of an incentive’ is used to refer to the strength of participants’ reactions to an incentive, whereas ‘sustainability’ refers to the persistency of this reaction, while efficiency of an incentive refers to the summarised costs and benefits for all affected stakeholders (Balliet et al., 2011).

Incentives differ in several ways, making them different in their effectiveness, sustainability and efficiency. One differentiation is based on whether there is a reward for desired behaviour or a punishment for undesired behaviour. Other major differences are the probability of occurrence when behaviour is desired or undesired and the magnitude of the incentive or penalty. Balliet et al. (2011) lists further possible differentiations for incentive designs, e.g., costs for giving incentives, centralised versus decentralised source of incentives, matching procedures, iterations, type of dilemma and cost-to-fine ratio. Sutter et al. (2010) find that incentive mechanisms increase cooperation more when they are chosen internally by group members compared to externally determined incentive mechanisms. Moreover, Herrmann et al. (2008) show that the cultural context also matters for the incentive efficiency.

Ex ante tests are an opportunity for testing policy measures in advance at low costs. For example, Viceisza (2012) emphasises the relevance of field experiments for policy recommendations especially for developing countries. Such experiments help to find and implement the best design for a policy measure, increasing the effectiveness, sustainability and efficiency of such policy measures. So far, most ex ante policy recommendations are based on the rational choice approach, assuming a homo-oeconomicus approach (Veetil, 2011). Since this assumption has been recently challenged in the field of behavioural economics, it seems necessary to account for insights found in the lab (Veetil, 2011). In this context, Levitt and List (2007) and a later contribution by Camerer (2011) discuss generalisability of lab experiments to the field. Among others, they discuss the importance of contextual features when determining the generalisability. Consequently, a framing that includes these contextual features is a prerequisite for enabling reasonable ex ante policy recommendations. A business simulation game provides a means for generating such a framing which can include selection of participants in a framed field experiment, generating a relatively high external validity (Harrison and List, 2004). Although the usefulness of ex ante testing of policy measures is undisputed, the promise of adaptation for contextual features with a business simulation game has seldom been used.

The objective of this paper is an ex ante testing of different policy measures on the reduction of fertiliser use for the case of small-scale palm oil producers on Indonesia's island of Sumatra. We use a multi-period business simulation game applied within a framed field experiment. We test incentives differing in their design, i.e., reward or punishment with different magnitudes or probabilities of occurrence but with constant effects on expected income. Thus the effectiveness, sustainability and efficiency of these incentives can be compared. Our research focuses on the Jambi province, a region with a very dynamic development in palm oil production (Laumonier et al., 2010; Wilcove and Koh, 2010). By doing so, we target the current and probable future small-scale palm oil producers that might be affected by such incentives. We are the first to test differently designed policy measures by means of a business simulation game in a framed field experiment. As a result, we hope to create a framing which enables adapting for contextual features in a way that allows generalising the results for policy recommendation (Camerer, 2011; Levitt and List, 2007).

The remainder of the paper is structured as follows: In section 2, the literature is discussed, and hypotheses are generated. In section 3, the method and experimental design is illustrated. Section 4 gives a description of the sample selection and data, while section 5 explains the data analysis. Section 6 presents the results along with discussion. Section 7 concludes.

2. Literature and hypothesis generation

In order to internalise the negative externalities through palm oil production, an appropriate incentive system is required. The design of such an incentive can differ inter alia in its probability of occurrence, its magnitude or in being a reward or a punishment. Problems with enforcing laws concerning probabilities and magnitudes of sanctions have been discussed theoretically by Polinsky and Shavell (1999). In combination, magnitude and probability of occurrence influence the effect on the expected income, i.e., participants' average costs or benefits for compliance or non-compliance behaviour. In theory, as long as the effects on expected income are held constant, a perfect rational, profit maximising agent will behave independently of the probability and the magnitude of an incentive (Becker, 1974). This would imply that for such an agent, it makes no difference on the behaviour if the incentive is designed e.g. as a reward or punishment, as long as the effects on expected income are held constant. Nevertheless, these assumptions have recently been challenged in the field of behavioural economics (Veetil, 2011). The literature on experimental economics shows that the reactions towards an incentive do not only depend on the effects on expected income. It is found that the incentive design matters, e.g. rewarding or punishing, the probability of occurrence or the magnitude of an incentive (Balliet et al., 2011). Subsequently, we generate our hypotheses, followed by a discussion of our expected results based on the corresponding literature.

In this paper, we ex ante analyse the effectiveness, sustainability and efficiency of differently designed incentives for policy measures. Hence, we vary the incentive design while holding the effects on expected income constant. Therefore, the first hypothesis, i.e. *"H1: For the same effect on expected income, the effectiveness of an incentive is independent of its design, i.e., reward or punishment, with different magnitudes and probabilities of occurrence"*, concerns effectiveness, which means the strength of participants' reactions to differently designed incentives. The second hypothesis, i.e. *"H2: For the same effect on expected income, the sustainability of an incentive is independent of its design"*, focuses on the sustainability, which means the persistency of the reaction through the incentive. The third hypothesis is *"H3: For the same effect on expected income, the efficiency of an incentive is independent of its design."* Here, efficiency means that the summarised stakeholders' costs for reducing the use of fertiliser are as low as possible. To discuss the expected results of the hypotheses, we employ Figure 1.

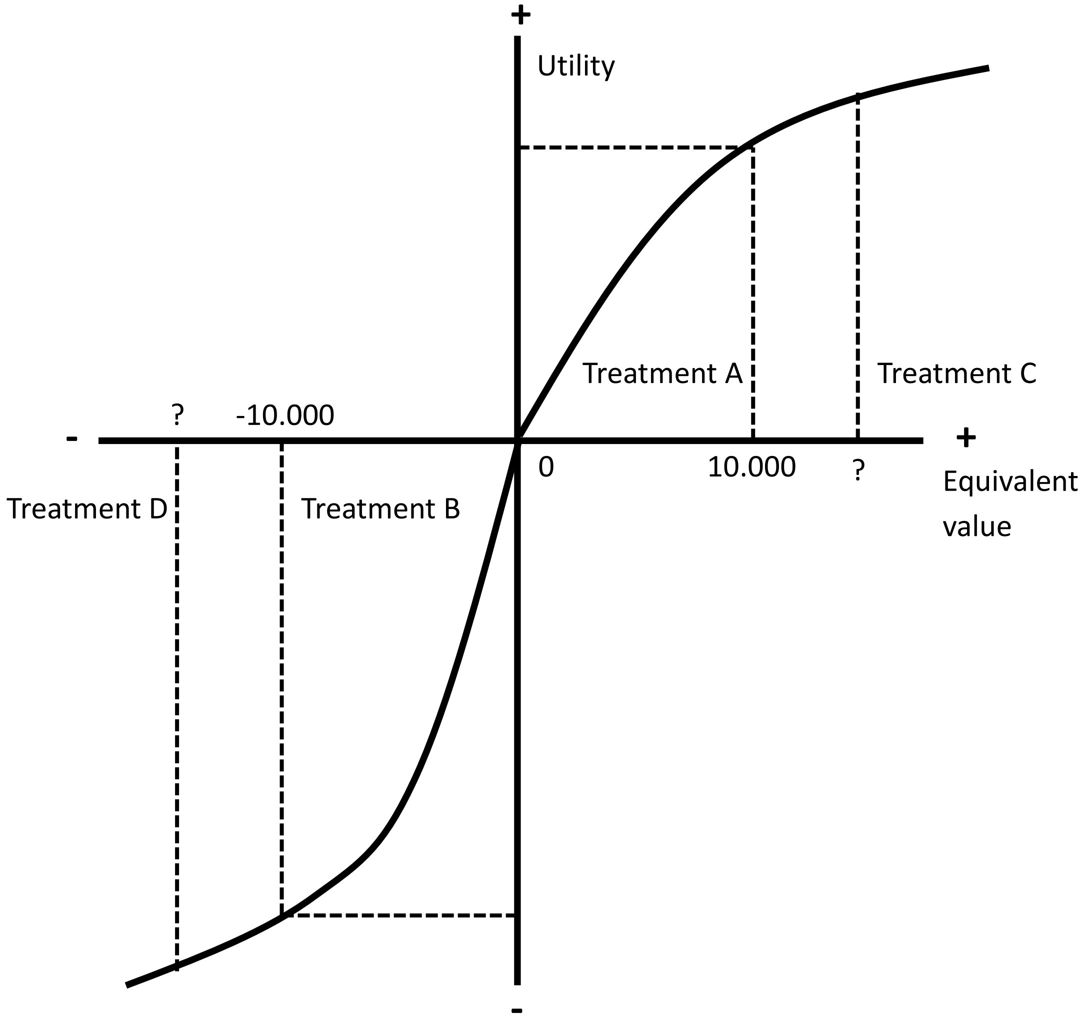


Figure 1. Expected utility effects through equivalent values
Source: Kahneman (2011), Author's own additions

Figure 1 provides an overview of the utility effects on the y-axis caused by the equivalent value of gains and losses presented on the x-axis. According to our definition, the equivalent value shows how participants value a certain treatment, which does not necessarily equal the expected value of this treatment. Four different treatments are charted in Figure 1, i.e., treatment A, treatment B, treatment C and treatment D with an equivalent value of 10,000, -10,000, more than 10,000 and less than -10,000, respectively. For this paper, a treatment is considered a reward when it has a positive equivalent value, whereas with a negative equivalent value it is a punishment.

For the consequences of gains or losses, a loss aversion is often found in the literature. For example, Heath et al. (1999) found that people are more motivated to avoid losses than to gain profits, with the precondition that there is a reference point which determines losses. Kahneman (2011) gives further examples for loss aversion, e.g., for cab drivers in New York

(Camerer et al., 1997), for Swiss messenger services (Fehr and Goette, 2007) or for professional golf players (Pope and Schweitzer, 2011). According to these papers, punishing should be more effective than rewarding. However, a meta-analysis involving 187 effect sizes from Balliet et al. (2011) shows reward and punishment having similar positive effects on cooperation, even though differences occur depending on the context. A framed field experiment from Ibanez and Martinsson (2008), which combines reward and punishment, shows that coca farmers react more strongly to increasing relative profits than to an increasing probability of occurrence for punishment. To summarise, there is evidence that the effectiveness of reward and punishment differs according to context, but they do not differ in general. Comparing gain and loss equals comparing treatment A and treatment B in Figure 1. The loss aversion leads to a bigger effect on the utility for treatment B compared to treatment A. However, in this study we test for a loss aversion in hypothesis one, i.e. participants react stronger to a certain punishment than to a certain reward.

For losses occurring with low probability, Kahneman (2011) state that a low loss probability leads to risk-avoiding behaviour. Anderson and Stafford (2003) found that for the same expected punishment, raising the magnitude causes more reaction than raising the probability of occurrence. Furthermore, Block and Vernon (1995) found that, for punishment, the behaviour depends on the participating group, but usually the reaction to the magnitude is stronger than the reaction to its probability of occurrence. These results imply that, for example, a punishment of 100,000 Rp with a 10% probability would probably have a higher impact than a certain punishment of 10,000 Rp. This means that the first punishment has a higher utility effect than the second one, holding constant the effects on expected income. Transferred to Figure 1, this is shown by the certain loss of treatment B compared to the uncertain loss of treatment D. In the context of our hypothesis one, we therefore expect an uncertain loss to be more effective than a certain one, holding constant the effects on expected income.

The literature on comparing rewards with a low probability of occurrence to certain rewards is very limited. In this context, Kahneman (2011) claims risk-loving behaviour is associated with rewards with low probability. Furthermore, Volpp et al. (2008) gave incentives for reducing weight for overweight individuals. They found that people react significantly to the incentives but not significantly differently between reward with a high and uncertain magnitude or a low and certain reward. In summary, we expect a stronger reaction to high and uncertain rewards compared to certain rewards, holding constant the expected value. This implies that an uncertain reward would create more impact than a certain reward, resulting in a stronger effect on the utility. This is represented with treatment C and treatment A in Figure 1, respectively. Hence, for hypotheses one, we expect that an uncertain reward is more effective than a certain one.

The expected results for H2 regarding the sustainability of incentives are based on repeated experiments. For example, Anderson and Stafford (2003) found that punishment has negative rather than positive effects on compliance for repeated public good games. On the other hand, in their meta-analysis, Balliet et al. (2011) discuss the sustainability of incentives, i.e. the persistency of an incentive effectiveness, on cooperation. They found that incentives are more effective for iterated experiments in general but do not indicate whether reward or punishment is more sustainable. Moreover, they do not discuss how the magnitude or the probabilities of incentives affect its sustainability. For our H2, our null hypothesis is no difference in the sustainability of different incentive designs.

The influence of incentive design on efficiency is another dimension in this analysis. In the literature, a public good game is often applied to find out more about the efficiencies of incentives. In their meta-analysis, Balliet et al. (2011) as well as Fehr und Gächter (2000) and Gächter et al. (2008) show that punishment generates additional gains, but the efficiency of reward and punishment is not compared. However, Rand et al. (2009) and Sefton et al. (2007) show that reward is more efficient than punishment. Nevertheless, little is known about how efficiency develops if the probability of occurrence of these incentives varies, holding constant expected values. With H3, we further investigate in this aspect and expect rewards to be more efficient.

3. Method and experimental design

To test these hypotheses, a business simulation game within framed field experiment is conducted. Framed field experiments are defined as conventional lab experiments but with a nonstandard subject pool as well as with field context (Harrison and List, 2004). In this way, for example, one possible bias is prevented by considering the cultural context (Balliet et al., 2011; Herrmann et al., 2008). A business simulation game simultaneously allows for creating a framing relatively close to reality. This allows for generalising the results found in the experiment to the field, which is especially important for assessing ex ante policy measures. Additionally, to test participants' ability to solve abstract problems, a quiz was conducted before the experiment started (Ihli and Musshoff, 2013).

To avoid confusion among participants, it is emphasised that some parts of the business simulation game are strongly simplifications of reality. The situation with which participants are confronted in the business simulation game is explained to them as follows: Each participant is supposed to manage a one-hectare palm oil plantation that is already established and producing. The farmer has to do so for 10 rounds, where each round is an equivalent for one year. At the beginning of each round, the participant has to decide how much fertiliser to use. It is assumed that the output on the simulated plantation depends solely on the amount of fertiliser used in the corresponding round, i.e., weather-related output fluctuations, emerging diseases, etc. are not taken into consideration. Each participant gets the same output for the same amount of fertiliser. To generate profit, the entire harvest is sold at the end of the round at a randomly determined price. Since palm oil bunches have to be processed promptly after harvesting because they will otherwise spoil, this is somewhat realistic. Moreover, the participants were told that there is a policy measure in the form of a subsidy to raise farmers' incomes and that this policy measure can change at any time during the experiment. If necessary, liquidity is always available, and there is a zero interest rate for the gained profit as well as for credits. After giving participants an overview of the experiment, we explained the procedure in greater detail.

Participants were not told that the change in policy measure always occurs after the fifth round. However, for reasons of improved readability, rounds 1-5 are called the 'first sequence' of the experiment, whereas rounds 6-10 are called the 'second sequence'. To stimulate a realistic behaviour, participants were given a real, monetary incentive. Thus, the gained profits from the ten rounds are added up, and 5 per cent of these profits is paid to them in the form of a shopping voucher. The procedure for each round in the business simulation game is constant: The first step is the participant's decision of how much fertiliser to use. In the second step, for each farmer individually, a random process fixes the price for the output for the corresponding round. In the third step, an enumerator determines the achieved output and price and calculates the profit.

In more detail, the first step in the business simulation game is the participant's decision of how much fertiliser to use in the current round. This is the only decision a participant has to make in each round. At that moment, farmers do not yet know the product price after harvest. The total amount of fertiliser used solely determines the output of the simulated oil palm plantation. A quadratic production function is applied to calculate the output. The functional form is $f(x) = 10 + 0.05x - 0.000055x^2$, where $f(x)$ and x represent the output amount and fertiliser, respectively. To reduce the mental effort for the participants, only 10 kg increments for fertiliser were allowed. The corresponding outputs are transferred into a table and a graph (Figure A1), which is handed out to the farmers at the beginning of the experiment. The equation behind this table was not shown and explained to the farmers. The table starts with 0 kg fertiliser and ends with 590 kg fertiliser. Although participants were allowed to use amounts of fertiliser beyond this scale, this happened rather seldom since the maximum output is achieved with 440 kg of fertiliser. A report by the FAO (2005) recommends a similar amount of fertiliser being used for palm oil plantations. In our business simulation game the possible outputs of palm oil bunches range from 10 tons to a maximum of 21.4 tons, which corresponds to the approximate annual output per hectare found in the literature (Fairhurst and McLaughlin, 2009; Jelsma et al., 2009).

The second step in the experiment is the determination of the price. This is evaluated once the farmer has decided on their use of fertiliser. To achieve a distribution of prices between participants, prices were determined separately for each person. A formula for calculating prices for the farmer is applied (Jelsma et al., 2009). The crude palm oil price strongly drives the bunch price that the farmers receive $Yield\ price\ for\ the\ farmer(\$) = Crude\ oil\ price(\$) \cdot 0.176 + \$10.2$. The average crude oil price from January to August 2012, which is the time prior to the start of the experiment, is about \$1,022/ton (Indexmundi, 2014)[2]. By assuming an exchange rate of 9,505 Rp/\$ (Xe.com, 2014) and holding everything else in the formula constant, we come to a price of 1,806,615 Rp/ton[3] for the farmers. To make numbers in the experiment more manageable, we divide all prices, incentives and penalties by 1,000. Therefore, we began with a rounded output price of 1,800 Rp/ton in the experiment. Since we mentioned this scaling to the farmers, we are sure that farmers can handle it without difficulties. For all following rounds, the price rises or falls by 200 Rp/ton of output based on the price of the previous round. This means that if the price of the previous round was, e.g., 1,800 Rp, the price either rises to 2,000 Rp or falls to 1,600 Rp with a 50% probability, respectively. Thus, the price development follows an arithmetic Brownian motion (Poitras, 1998) until the 10th round, with a minimum of 0 Rp and a maximum of 3,600 Rp. According to our above-mentioned calculation, bunch prices reached from 727 Rp/ton in August 2005 to 2,051 Rp/ton in February 2011. Since the chance to hit the lowest price of 727 Rp/ton is less than 1% we are realistic when assuming low prices. However, since farmers have experienced

[2] $Average\ OP(\$) = \frac{\$1{,}021+\$1{,}048+\$1{,}106+\$1{,}157+\$1{,}031+\$928+\$953+\$931}{8}$

[3] $BP(\$) = \$1{,}022 \cdot 0.176 + \$10.2$ and $BP(Rp) = \$190.07 \cdot 9{,}505 \frac{Rp}{\$}$

volatile prices in recent years, we assumed that they can also handle prices which are beyond 2,051 Rp/ton. Therefore, we think that such prices are not exaggerated in the farmers' perception. To evaluate whether the price rises or falls, participants drew a ball from a bag which contains three green and three red balls. Prices rise for drawing a green ball and fall for drawing a red ball.

The third step in each round is to calculate the realised profit. The revenue in a particular round equals the achieved output multiplied by the product price determined for the corresponding round. Since there is no option for storage, all produce has to be sold. The cost for fertiliser is fixed at 10,000 Rp/kg. To keep the right relation to the price, this cost is scaled to 10 Rp/kg, with profit being the revenue minus the cost of fertiliser used. The additional policy incentives and penalties are described below. After the profit has been calculated and the participants have been informed accordingly, the next round starts, and the participants must decide again on how much fertiliser they want to use. The rounds were done simultaneously for all participants. After 10 rounds, the experiment is over.

The policy measures differ between the first and the second sequence, except for the control treatment. For the first sequence, the participants always receive a fixed subsidy of 10,000,000 Rp per round, scaled to 10,000 Rp per round, without regard to their use of fertiliser. For the second sequence, this public subsidy can change to one out of four policy measures for reducing the use of fertiliser; during the second sequence this measure remains constant. At the beginning of the first round, the farmers are informed that subsidies may change once during the experiment but are not given any further information. At the beginning of the 6th round, each participant is treated with one of the five following policy measures, which are described in Table 1. These treatments for the second sequence are in line with treatments in Figure 1.

Table 1
Overview of the policy measures

Round[a)]	Treatment	Subsidy (Rp)	Fertilizer (kg) trigger	Reward[b)] (Rp)	Punishment (Rp)
1-5	Control, A, B, C, D	10,000	-	0	0
6-10	Control	10,000	-	0	0
6-10	A	0	≤ 120	10,000 (100%)	0
6-10	B	10,000	> 120	0	- 10,000 (100%)
6-10	C	0	≤ 120	100,000[b)] (10%)	0
6-10	D	10,000	> 120	0	- 100,000[c)] (10%)

Note: a) Round 1-5 and 6-10 are the first and second sequence, respectively; b) During the experiment's explanation to the participants, the reward was named 'compensation payment' c) To keep the incentives effect on expected income constant at the magnitude of 10,000 Rp, the incentives magnitude is adapted to its probability

It is noteworthy that for treatments A, B, C and D the effect on expected income is always 10,000 Rp if they use 120 kg fertiliser or less and zero if they use more than 120 kg fertiliser. Therefore, the effect on expected income is equal for all treatments, except for the control treatment. This means that a perfectly rational and profit maximising participant would behave independently of the treatment (Becker, 1974; Polinsky and Shavell, 1999). The magnitude of 10,000 Rp as well as the trigger of 120 kg are chosen arbitrarily at a level where the majority, but not all, rational, profit maximising participants would follow the incentive. For treatments C and D, the incentive or penalty only applies with a 10% probability, which is determined by each participant making a draw at the end of the round after the decision on fertiliser application had been made. Table A1 in the on-line appendix shows the profit-maximising use of fertiliser for a perfect rational participant in this experimental design. More detailed information about how the decision sheets for the experiment were designed is also presented in the on-line appendix (Figure A2 and Figure A3.)

4. Sample selection and data

The experiment was executed in the Jambi province in Sumatra, Indonesia. Jambi has about three million inhabitants and has an area of approximately 50,000 km². The research area is in four regions of the Jambi province, i.e., Tebo, Bungo, Batang Hari and Muaro Jambi. Due to the flat land, these regions are especially valuable for palm oil production. Over the last several decades, there has been a strong transformation of this landscape towards oil palm plantations (Laumonier et al., 2010; Wilcove and Koh, 2010), which will likely continue in the future, making it a useful research area for our experiment.

The data was collected from 29 randomly chosen villages from October to December 2012. For each village, an entire list of farmers was created. Depending on the size of the village in terms of inhabitants, between 10 and 18 small-scale farmers were chosen randomly and invited to participate in the experiment. The experiment started in the afternoon or in the evening after prayer time to accommodate to local customs. The experiment was completed in available public rooms or sometimes in the house of the head of the village. The business simulation game was done simultaneously for the whole group. Before the experiment started, participants were asked about their socioeconomic data. To avoid conflicts, every participant from a village received the same treatment. At the beginning, each participant received a questionnaire including one version of the business simulation game. Then an enumerator explained the procedure with the support of visual posters, and the participants had the chance to ask questions. During the experiment, the participants were divided into subgroups, so each enumerator was monitoring between 3 and 5 participants. This structure enabled the participants to ask questions on a more personal level. This is especially important for cultural reasons, since participants often hesitate to ask when in a large group.

In total, 328 small-scale farmers participated in the experiment. Nine uncompleted questionnaires were dismissed, resulting in 319 participants for the analysis. On average, participants made a profit of 37,940 Rp per round, resulting in 18,970 Rp worth of shopping vouchers for the entirety of the business simulation game. Considering that the average daily wage for a worker is around 50,000 Rp in the research area, this seems to be a sufficient compensation for participating in this one-hour experiment.

Table 2
Average socioeconomic data and fertiliser use

	Control Treatment	Treatment A	Treatment B	Treatment C	Treatment D	Kruskal-Wallis Test[a)]	N[b)]
Gender[c)]	0.81	0.86	0.83	0.93	0.94	0.111	319
	(0.05)	(0.04)	(0.05)	(0.03)	(0.03)		
Age	44.2	44.5	40.9	42	44.4	0.112	319
	(1.5)	(1.3)	(1.4)	(1.6)	(1.4)		
Education, years	7.6	7.6	8.6	7.4	7.7	0.340	319
	(0.4)	(0.4)	(0.5)	(0.4)	(0.4)		
Household size	4.4	4.5	4.1	4.4	4.4	0.566	319
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)		
Land owned in hectare	2.2	1.9	2.7	3	3.8	0.000	319
	(0.2)	(0.3)	(0.3)	(0.3)	(0.7)		
Main fruit oil palm [d)]	0.16	0.11	0.03	0.43	0.33	0.000	319
	(0.04)	(0.04)	(0.02)	(0.06)	(0.06)		
Quiz	4.4	4.5	4.5	4.3	4.4	0.908	319
	(0.2)	(0.1)	(0.2)	(0.2)	(0.2)		
Fertiliser, round 1-5[e)]	276	286	275	254	242	0.002	1,595
	(8)	(8)	(10)	(9)	(9)		
Fertiliser, round 6-10[e)]	285	176	169	171	158	0.000	1,595
	(8)	(8)	(8)	(8)	(6)		
Price	1825	1878	1825	1847	1773	0.000	3,190
	(14)	(16)	(16)	(16)	(15)		
Participants	70	65	60	61	63		319

Note: a) p-values; b) Observations; c) 1=male, 0=female, d) 1=oil palm, 0=others, e) Average amount kg fertiliser per treatment, standard deviation in parentheses

Table 2 shows the socioeconomic and experimentally collected data of the participants divided among the treatments. For six participants we lack information about the household size. In these cases, we assumed values of 4, since with a share of over 40% this is the most frequent value in our dataset. For five other participants we lack information about the owned land. Here, we assume a value of zero land ownership.

For the quiz, which tested participants' ability to solve abstract problems, a Kruskal-Wallis test shows no significant differences between the treatments. Also, for gender, age, years education and household size no difference between the treatments was found. For the variables "owned land" and "main fruit oil palm", we found significant differences. In the Jambi province, the concentration of oil palm plantations decreases with the distance from the capital Jambi City. Also, the average owned land size varies significantly between the villages. We have 29 villages in the dataset, resulting in approximately 6 villages per treatment. This might be too few to balance these inequalities, which could explain the differences between the treatments for owned land and main fruit.

Surprisingly, we also found significant differences in the average price per treatment. As described above, the generating of the price was designed to be random for each participant. During the experiments, one of the authors was always present to guarantee a proper conduction of the experiment and we have no proper explanation for what could cause these differences in prices per treatment, other than that 9 rounds of experimentally adjusted prices is hardly sufficient to reveal the genuinely random nature of the pricing process used here. Figure A4 in the appendix gives a better overview over the prices per treatment.

In the first sequence of the business simulation game, participants use 267 kg of fertiliser on average. For the second sequence, the control treatment uses 285 kg, whereas the four treatment groups use 168 kg of fertiliser on average showing that participants react to both incentives and penalties for appropriate use. Since the treatment for the first sequence is equal for each participant, it was expected that treatment groups do not differ in their use of fertiliser. A Kruskal-Wallis test shows that at the 1% level this expectation does not hold.

For each participant, the use of fertiliser can differ in two ways, i.e. by the average level of fertiliser use and by the learning effect which can occur during the rounds. We estimate a regression with the initial fertiliser use (rounds 1 – 5) as dependent variable, while using the individual socioeconomic data and dummy variables for each village as explanatory variables. To estimate the learning effect, we use the same variables multiplied with the number of the round in which the observation occurred.

Table 3
Test for village-level effects

	Coefficient	SE	p-value[a)]
Level: quiz	9.99	7.4	0.18
Level: gender[b)]	-72.48	27.5	0.01**
Level: age	-0.74	0.9	0.42
Level: education, years	3.19	3.4	0.34
Level: household size	-7.05	5.8	0.22
Level: land owned, in hectare	2.24	2.9	0.44
Level: oil palm, main fruit	56.07	37.6	0.14
Level: village 1	-0.56	78.5	0.99
…	…	…	…
Level: village 29	17.92	80.4	0.82
Learn: quiz	0.65	2.2	0.77
Learn: gender[b)]	7.07	8.3	0.39
Learn: age	-0.17	0.3	0.54
Learn: education, years	-0.04	1.0	0.97
Learn: household size	1.48	1.7	0.40
Learn: land owned, in hectare	-0.12	0.9	0.89
Learn: oil palm, main fruit	-5.11	11.4	0.65
Learn: village 1	8.11	25.5	0.75
…	…	…	…
Learn: village 29	15.81	23.7	0.50
Price	0.01	0.0	0.68
Constant	259.92	98.2	0.0**
Observations	1,595		
Adjusted R^2	0.22		

Note: a) * p < 0.05; ** p< 0.01; *** p< 0.001; b) 1=male, 0=female

Table 3 shows the results of this estimation. A likelihood-ratio test, comparing this regression with and without socioeconomic data, clearly indicates that for this analysis socioeconomic data improve the model. For the level effects, one village and gender are the only variables with a coefficient significantly different from zero at the 5% level. The coefficient for gender indicates that men tend to use less fertiliser than woman. Moreover, an F-test shows that the 29 village-level effects are not equal at the 5% level. It seems that there are unobserved village-specific effects which influence the use of fertiliser. These village-specific effects may explain the differences between the treatments shown in Table 2.

For the learning effect, only one village and none of the socioeconomic variables show a significant influence on the use of fertiliser. Moreover, an F-test does not indicate that the 29 village learning effects are not equal at the 5% level. Surprisingly, the price also shows no significant influence on the use of fertiliser in this regression.

5. Data analysis

This section discusses the methods of data analysis. To begin with, the necessity of a matching procedure for a meaningful analysis is explained, followed by a description of this procedure. We then turn to the estimation of the incentives' effectiveness, sustainability and efficiency.

5.1. Matching procedure

In this analysis, the first sequence is the baseline for estimating the incentives' effectiveness, sustainability and efficiency. To ensure comparability, this baseline has to be equally independent of the incentive design, i.e. treatment. Therefore, the treatment differences in fertiliser use in the first sequence, shown in Table 2, have consequences for the analysis. For example, in the fifth round, participants in treatment D have an average fertiliser use of about 243 kg, whereas participants in treatment A have an average use of 330 kg fertiliser. The trigger for the incentive in the second sequence is always 120 kg fertiliser. Therefore, treatment D participants have to reduce their use by 123 kg, whereas treatment A participants have to reduce their use by 210 kg to pass the trigger level of 120 kg of fertiliser. Since we investigate the changes of behaviour through different designed incentives, and the baseline for our analysis is the first sequence, we cannot compare the incentives with this difference of fertiliser use in the first sequence. Table 3 shows little influence of prices on the use of fertiliser, indicating that prices are not a source that might explain the differences in the use of fertiliser.

To overcome the problem of different fertiliser use among the treatments in the first sequence, i.e., rounds one to five, a minimum Euclidian distance matching is applied. This matching method, which is widely used in the literature (Tiedemann and Latacz-Lohmann, 2013), is intuitive and allows for matching the fertiliser use in the first sequence. The idea behind applying this method is as follows: If participants of the treatments behave similarly for rounds one through five, and then they behave differently for rounds six through ten, this difference is caused by the difference in the treatment. To generate an equal basis for all treatments in the first sequence, treatments A, B, C and D are each matched separately with the control treatment, resulting in four independent matching procedures. For each of these matching procedures, one participant of the respective treatment is chosen for each participant from the control treatment. The Euclidian distance between the participant of control treatment i and the participant of the respective matched treatment j is calculated as follows:

$$d_{ij} = \sum_{k=1}^{5} \left(x_{ik} - x_{jk}\right)^2 \tag{1}$$

In equation (1), d_{ij} is the Euclidian distance between participant *i* from the control treatment and participant *j* from the respective matched treatment, i.e., either treatment A, B, C or D. The variable *k* represents the round, ranging from one to five, while the variable *x* is the amount of fertiliser used. Thus, the Euclidian distance is the squared difference in fertiliser use between two farmers, which are summed up from rounds one to five. For each participant from the control group, the participant *j* from the matched treatment with the minimum Euclidian distance is used for matching; this equals a nearest neighbour matching. To get a proper trade-off between variance and bias for this matching procedure (Caliendo, 2006), each participant from treatment *j* can be repeatedly taken for the matching to a maximum of three times. A more frequent use of one individual participant would automatically exclude the consideration of other participants, leading to greater loss of information, whereas a less frequent use would lead to an insufficient matching result. In the special case that there are two participants with the same Euclidian distance, both of them are included for the further analysis. This leads to 70, 71, 75, 78 and 81 observations used for data analysis for the control treatment, treatment A, B, C and D, respectively. As expected, after the matching, a Kruskal-Wallis test indicates no significant difference in the level of use of fertiliser at a p-value of 5 % between the treatments in the first sequence. Thus, after the matching we can use the first sequence as a baseline for the analysis without restriction.

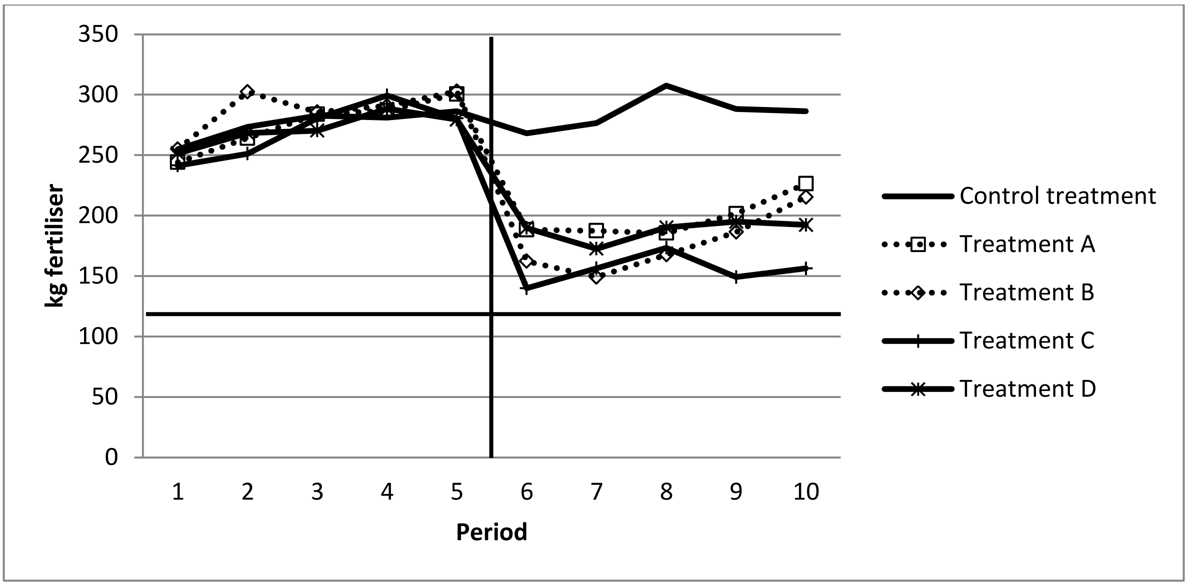


Figure 2. Average amount of fertiliser used per round and treatment after matching
Source: Author's own illustration

Figure 2 shows the mean amount of fertiliser used per round and per treatment after the matching procedure. The horizontal line presents the 120 kg fertiliser trigger for the incentives, whereas the vertical line stands for the policy change after round 5. A look at Figure 2 already indicates differences between the treatments for the second sequence.

5.2. Generating variables and regression analysis

The matched data are the basis for evaluating the differences in the treatments. To test for our hypotheses, we compare sequence one, where each participant is treated equally, with sequence two, where the different treatments take place. The effectiveness of the incentives is defined as the different levels of fertiliser use between sequence one and sequence two.

Figure 2 indicates that the average amount of fertiliser used rises within the first sequence. Most of the participants are below the profit maximum level of around 400 kg fertiliser use. Assuming that maximising profit influences the farmers' decisions, this increasing use of fertiliser in the first sequence could reflect a learning effect. To account for that effect in the first sequence, one learning variable is used for all treatments, from zero to 4 for rounds 1 to 5 and held constant at 4 throughout the second sequence.

For the second sequence the 120 kg trigger for the incentives discourages fertiliser use above this level. For these treatments, rounds 6 to 10 show whether fertiliser use remains stable at the decreased level caused by the incentive. If so, this development is seen as an indicator for sustainability. The sustainable variables are generated separately for each treatment. It starts with one in round six and then increases by one for each following round. For the control treatment, no incentive restricts the use of fertiliser, so the learning effect influenced by the profit maximum could go on. For a better comparability with the sustainability variable, the variable for this learning effect is generated starting with one in round six and then increasing by one for each following round. Moreover, the learning effect is also controlled for socioeconomic data. To generate the necessary variables, the value of the specific socioeconomic variable is multiplied with the number of the round in which the observation is made. By doing so, we correct for influences caused by different farmer ages, gender, education, household size, owned land (hectare) and the ability to solve abstract problems measured with a quiz.

Our producer efficiency measure for each treatment is taken as the profit difference between the profit maximum when farmers are compliant and follow the incentive, at the trigger of 120 kg fertiliser use, and the profit maximum when participants are non-compliant and ignore the incentive, which is usually around 400 kg fertiliser use. This measure shows the costs of reducing the use of fertiliser in terms of reduced profit for the farmers. Inclusion of this variable for the second sequence reflects the differential profit motive in explaining participants reactions to the incentive or penalty treatments, and is calculated for each round of the second sequence. To avoid an influence on treatments shifts, we measure the profit differences relative to the mean value for each treatment.

A Hausman test indicates that a random effect model would lead to inconsistent results. Thus, with the availability of panel data, a fixed effects regression allows us to account for unobserved, unchanging participant-specific effects. Moreover, the fixed effects include the influence from observed village fixed effects or socioeconomic data on the level of fertiliser use, which is shown in Table 3.

$$y_{it} = a_i + \sum_{k=1}^{K} \alpha_k x_{kit} + u_{it} \tag{2}$$

Equation (2) gives the specification of the applied fixed effects model. y_{it} presents the applied fertiliser of farmer i in round t. a_i stands for the corresponding fixed effects for farmer i. α_k is the estimated coefficient for variable k, whereas x_{kit} is the value of variable k from farmer i in round t. u_{it} is the error term of the model.

6. Results and discussion

Table 4
Estimation results of fixed effects model

	Coefficient	SE	p-Value[a)]	
Shift, control treatment	24.8	31.9	0.44	
Shift, treatment A	-47.1	23.5	0.05	*
Shift, treatment B	-95.3	23.3	0.00	***
Shift, treatment C	-80.7	23.4	0.00	***
Shift, treatment D	-50.8	22.9	0.03	*
Learning, sequence 1, all	2.26	4.12	0.58	
Learning, sequence 2, control treatment	-2.61	5.18	0.61	
Sustainability, treatment A	-1.84	5.23	0.72	
Sustainability, treatment B	8.32	5.01	0.10	
Sustainability, treatment C	-5.80	5.04	0.25	
Sustainability, treatment D	-4.90	5.11	0.34	
Price, sequence 1, all	0.037	0.011	0.00	**
Price, sequence 2, control treatment	0.008	0.014	0.54	
Profit difference, treatment A	0.020	0.002	0.00	***
Profit difference, treatment B	0.010	0.002	0.00	***
Profit difference, treatment C	0.004	0.002	0.03	*
Profit difference, treatment D	-0.003	0.002	0.18	
Learn: quiz	0.24	0.46	0.60	
Learn: gender	6.14	1.77	0.00	***
Learn: age	-0.01	0.05	0.91	
Learn: education, in years	-0.67	0.19	0.00	***
Learn: household size	1.52	0.37	0.00	***
Learn: land owned, in hectare	-0.22	0.13	0.09	
Learn: oil palm, main fruit	3.12	1.33	0.02	*
Constant	183.25	20.92	0.00	***
Observations	3,750			
Adjusted R^2	0.63			

Note: a) * $p < 0.05$; ** $p< 0.01$; *** $p< 0.001$

Table 4 shows the results from estimating the fixed effects model shown in equation (2). Except for the control treatment, all treatments show a shift in the second sequence at the 5% significance level. This indicates that each incentive significantly decreases the use of fertiliser. Treatment A and treatment D with 47.1 kg and 50.8 kg, respectively, show the lowest shifts. Wald tests clearly indicate that treatment B and treatment C with 95.3 kg and 80.7 kg, respectively, have significantly greater reduction at a 5% level. Within these two pairs of treatments, no significant difference is found. Treatment B and treatment C are significantly more effective implying that certain punishment and uncertain reward are more effective than other treatments, holding effects on expected income constant. This means that our first hypothesis, i.e. *"H1: For the same effect on expected income, the effectiveness of an incentive is independent of its design"*, is rejected.

Comparing certain reward (treatment A) with certain punishment (treatment B), we find punishing to be more effective. This is in line with the literature (Heath et al., 1999; Camerer et al., 1997; Fehr and Goette, 2007; Pope and Schweitzer, 2011). Also, in their meta-analysis regarding cooperation in social dilemmas, Balliet et al. (2011) found slightly higher but not significant differences in the effectiveness of punishment compared to reward. The result that uncertain and high rewards (treatment C) lead to stronger reactions than certain rewards (treatment A) is consistent with some literature (Kahneman, 2011). Participants seem to prefer a risky reward to a secure.

For punishment, we found the certain punishment (treatment B) to be more effective than uncertain and severe punishment (treatment D), implying that participants prefer to take the risk of the uncertain punishment. This finding contradicts the literature where the opposite effect was found (Block and Vernon, 1995; Anderson and Stafford, 2003; Kahneman, 2011).

For sustainability, a positive coefficient means that the average use of fertiliser rises in the second sequence. This means that the achieved drop in the use of fertiliser through the incentive gets lost and the incentive is considered as being unsustainable. Except treatment B, any treatment has a sustainability coefficient which is significantly different from zero. In case of treatment B, the sustainability effect is significantly different from zero at a 10% level. Moreover, a Wald test indicates that treatment B's coefficient is significantly higher than for all the other treatments at a 5% level. Treatment B, i.e., a certain punishment, is not as sustainable as the other treatments. Therefore, our second hypothesis, i.e. *"H2: For the same effect on expected income, the sustainability of an incentive is independent of its design"*, is also rejected.

Sustainability can be compared with iterated dilemmas. In their meta-analysis, Balliet et al. (2011) found punishment to be insignificantly more sustainable than reward. On the other hand, Anderson und Stafford (2003) found that punishment has a rather negative than positive effect on compliance for repeated public good games. This is consistent with our results, since certain punishment (treatment B) is the only treatment that is significantly less sustainable than the other treatments. If we combine the results of hypothesis one and two, we find that for creating a high and sustainable reduction for fertiliser use, treatment C, i.e. high but uncertain reward, is the preferable design for an incentive.

The profit difference is our measurement for efficiency in this experiment. For one Rp in profit difference, participants use fertiliser at 0.020 kg, 0.010 kg, 0.004 kg and -0.003 kg for treatment A, treatment B, treatment C and treatment D, respectively. Since Wald tests indicate different coefficients between all of these treatments at 5% level, this order also represents the order of efficiency for the different treatments. For the certain incentives, i.e. treatment A and treatment B, as well as for reward incentives, i.e. treatment A and treatment C, participants show a significantly stronger reaction to the profit difference than they do to the uncertain or punishment incentives. Our third hypothesis, i.e. *"H3: For the same effect on expected income, the efficiency of an incentive is independent of its design"*, is also rejected.

Participants react more strongly to the profit difference for certain and for reward incentives. The later statement is in line with previous research that found reward being more efficient than punishment (Rand et al., 2009; Sefton et al., 2007). In these papers, punishing can have a crucial effect on lowering the efficiency by generating costs. In our experiments, no such costs occur, which underlines again our statement of reward being more efficient.

The learning effect in the first sequence is estimated simultaneously for all participants. The profit maximum would be reached with a fertiliser use of around 400 kg, depending on the price. Most of the participants are clearly below this level of fertiliser use; thus we can expect some adoption. The learning effect, i.e. the general trend in the amount of used fertiliser, is not significantly different from zero in the first sequence and is also not significantly different in the control treatment in the second sequence. For the socioeconomic data, the learning effects are not significantly different from zero for the quiz, the age of the participants and the owned land. However, we found a significant positive learning effect for women, for household size and for palm oil farmers, whereas education has a significant negative influence on this learning effect at a 5% level. Moreover, the owned land has a negative influence on fertiliser use at a 10% level. It seems that the learning effect is not a general trend for all participants but only for certain groups of participants. The coefficient for the price effect is 0.037 and significantly different from zero for the first sequence. Thus participants anticipate price developments of 7.5 kg for a price increase of 200 Rp. Interestingly, for the available control treatment in the second sequence, price effects are not significant.

7. Conclusion and outlook

The use of fertiliser in palm oil production creates negative externalities. If policymakers want to restrict such externalities, an effective, sustainable and efficient incentive is desirable. The aim of this paper is to test ex ante differently designed incentives on their effectiveness, sustainability and efficiency, while holding effects on expected income constant. We use a business simulation game within a framed field experiment to test for policy measures. This method might improve the generalisability of the results, which is important for reasonable policy recommendations. Also, reward and punishment with different probabilities and magnitudes are simultaneously tested for policy recommendation in a framed field experiment for the first time, as far as we know.

The hypotheses are *"For the same effect on expected income, the H1: effectiveness, H2: sustainability, H3: efficiency of an incentive is independent of its design, i.e. reward or punishment, with different magnitudes and probabilities of occurrence."* After testing for H1, we found that participants react more strongly to either certain punishment or uncertain reward, holding constant the effects on expected income. For punishment, we did not expect a stronger reaction for the certain design, whereas for reward we expected the stronger reaction towards the uncertain design. By testing H2, we found that the reaction towards uncertain reward is also sustainable, leading to the result that uncertain reward with high magnitude is the preferable design. We expect no differences in the sustainability depending on the incentives design. Furthermore, by testing for the third hypothesis it was found that certain and reward incentives are significantly more efficient than uncertain and punishment incentives. Thus, a certain reward is the most efficient design. For reward incentives, this was expected. To sum up, our findings suggest that that by adapting the incentive's design for a policy measure, its influence can be improved, even with constant effects on expected income.

Our results have implications for policymakers. In general, we found that reward incentives are preferable over punishment incentives. Moreover, it depends on the aim of a policy measure whether the reward should be certain or uncertain. If policymakers desire a significant reduction in fertiliser use, i.e., an effective and sustainable incentive, results suggest using an uncertain reward. However, if policy aims to reduce fertiliser use at low cost for all stakeholders, i.e. an efficient incentive, results suggest using a certain reward.

For each of these two policy measures, there are considerations or costs that might be relevant for practical application. An uncertain reward, i.e. the most effective and sustainable incentive design, could be seen as unfair (Fehr and Schmidt, 1999). This may undermine the acceptance of a policy measure or raise other issues, e.g. in the case that two farmers are reducing their use of fertiliser, but only one of them gets rewarded. Moreover, it might be challenging to implement a process that is safe against corruption when it comes to allocating such a reward. A certain reward, i.e. the most efficient incentive design, presupposes for controlling every fertiliser-reducing farmer. This could involve high monitoring costs, which would decrease the efficiency of such a design in contrast to uncertain designs.

It is not clear if our results are specific for our particular context or if they are valid in general. For example, it is possible that due to cultural differences, results may differ in other countries or even in other parts of Indonesia (Balliet et al., 2011; Herrmann et al., 2008). Moreover, we do not know whether our results are specific for palm oil production or if they would be valid for other production systems. It is also unclear if results are specific for fertiliser use or if they can be applied for other inputs or management practices. Consequently, experiments which expand the research in such directions would be beneficial. Another possible research direction is to relax the assumptions of the experiment. Thus, it would be possible to allow for several production systems, which would make the production decision more flexible. Additionally, one could apply randomly selected prices from the past instead of our synthetic method. Furthermore, by implementing monitoring costs for an incentive, particularly the efficiency measure could be different. However, the applied business simulation game is a flexible and powerful method for obtaining results which are useful for both policymakers and researchers.

References

Anderson, L.R. and Stafford, S.L. 'Punishment in a regulatory setting: experimental evidence from the VCM', *Journal of Regulatory Economics,* Vol. 24, (2003) pp. 91–110.

Balliet, D., Mulder, L. B. and Van Lange, P.A.M. 'Reward, punishment, and cooperation: A meta-analysis', *Psychological Bulletin,* Vol. 137, (2011) pp. 594–615.

Becker, G.S. 'Crime and punishment: An economic approach', *Essays in the Economics of Crime and Punishment,* UMI, (1974) pp. 1–54.

Block, M.K. and Vernon, G.E. 'Some Experimental Evidence on Differences between Student and Prisoner Reactions to Monetary Penalties and Risk', *The Journal of Legal Studies,* Vol. 24, (1995) pp. 123–138.

Caliendo, M. 'Microeconometric evaluation of labour market policies.' Vol. 568. Springer, (2006).

Camerer, C. 'The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List', *Working paper,* (2011).

Camerer, C., Babcock, L., Loewenstein, G. and Thaler, R. 'Labor Supply of New York City Cabdrivers: One Day at a Time', *The Quarterly Journal of Economics,* Vol. 112, (1997) pp. 407–441.

Fairhurst, T., and McLaughlin, D. 'Sustainalbe Oil Palm Development on Degraded Land in Kalimantan', *World Wildlife Fund,* Washington, DC, (2009).

FAO. 'Fertilizer use by crop in Indonesia', *Food and Agriculture Organization of the United Nations (FAO),* First version, Rome (2005).

FAOSTAT. Available at: http://faostat3.fao.org/faostat-gateway/go/to/download/Q/*/E, (2013) (last accessed 18 October 2013).

Fehr, E. and Gächter, S. 'Cooperation and punishment in public goods experiments', *American Economic Review,* Vol. 90, (2000) pp. 980–994.

Fehr, E. and Goette, L. 'Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment', *The American Economic Review,* Vol. 97, (2007) pp. 298–317.

Fehr, E., and Schmidt, K.M. 'A theory of fairness, competition, and cooperation.', *Quarterly Journal of Economics,* Vol. 114, (1999) pp. 817-868.

Fitzherbert, E.B., Struebig, M.J., Morel, A., Danielsen, F., Brühl, C.A., Donald, P.F. and Phalan, B. 'How will oil palm expansion affect biodiversity?', *Trends in ecology & evolution,* Vol. 23, (2008) pp. 538–545.

Gächter, S., Renner, E. and Sefton, M. 'The long-run benefits of punishment', *Science,* Vol. 322, (2008) pp. 1510.

Harrison, G.W. and List, J.A. 'Field experiments', *Journal of Economic Literature,* Vol. 42, (2004) pp. 1009–1055.

Heath, C., Larrick, R.P. and Wu, G. 'Goals as Reference Points', *Cognitive Psychology,* Vol. 38, (1999) pp. 79–109.

Herrmann, B., Thöni, C. and Gächter, S. 'Antisocial punishment across societies', *Science,* Vol. 319, (2008) pp. 1362–1367.

Hewitt, C.N., MacKenzie, A.R., Di Carlo, P., Di Marco, C. F., Dorsey, J. R., Evans, M. et al. 'Nitrogen management is essential to prevent tropical oil palm plantations from causing ground-level ozone pollution', *Proceedings of the National Academy of Sciences,* Vol. 106, (2009) pp. 18447–18451.

Ibanez, M. and Martinsson, P. 'Can we do policy recommendations from a framed field experiment? The case of coca cultivation in Colombia', *Working Papers in Economics 306*, University of Gothenburg. School of Business, Economics and Law, (2008).

Ihli, H.J. and Musshoff, O. 'Investment Behavior of Ugandan Smallholder Farmers: An Experimental Analysis', *GlobalFood Discussion Papers,* No. 21, (2013).

Indexmundi. Available at: http://www.indexmundi.com/commodities/?commodity=palm-oil&months=120, (2014) (last accessed 28 November 2014).

Jelsma, I., Giller, K. and Fairhust, T. 'Smallholder Oil Palm Production Systems in Indonesia: Lessons from the NESP Ophir Project', Plant Productin Systems, Plant Sciences Group. University Wageningen, (2009).

Kahneman, D. 'Thinking, Fast and Slow', New York: Farrar, Straus and Giroux, (2011).

Keller, M. and Matson, P.A. 'Biosphere-Atmosphere Exchange of Trace Gases in the Tropics: Evaluating the Effects of Land Use Changes', *Global Atmospheric-Biospheric Chemistry,* Vol. 48, (1994) pp. 103-117.

Koh, L.P. and Wilcove, D.S. 'Is oil palm agriculture really destroying tropical biodiversity?', *Conservation Letters,* Vol. 1, (2008) pp. 60–64.

Laumonier, Y., Uryu, Y., Stüwe, M., Budiman, A., Setiabudi, B. and Hadian, O. 'Eco-floristic sectors and deforestation threats in Sumatra: identifying new conservation area network priorities for ecosystem-based land use planning', *Biodiversity and conservation,* Vol. 19, (2010) pp. 1153–1174.

Levitt, S. and List, J. 'What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?', *The Journal of Economic Perspectives,* Vol. 21, (2007) pp. 153–174.

Palm, C. A., Alegre, J.C., Arevalo, L., Mutuo, P.K., Mosier, A.R. and Coe, R. 'Nitrous oxide and methane fluxes in six different land use systems in the Peruvian Amazon', *Global Biogeochemical Cycles,* Vol. 16, (2002) pp. 1–21.

Poitras, G. 'Spread options, exchange options, and arithmetic Brownian motion', *Journal of Futures Markets,* Vol. 18, (1998) pp. 487–517.

Polinsky, A.M. and Shavell, S. 'The economic theory of public enforcement of law', *Working paper 6993,* National Bureau of Economic Research. Cambridge, (1999).

Pope, D.G. and Schweitzer, M. 'Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes', *The American Economic Review,* Vol. 101, (2011) pp. 129–157.

Rand, D.G., Dreber, A., Ellingsen, T., Fudenberg, D. and Nowak, M.A. 'Positive interactions promote public cooperation', *Science,* Vol. 325, (2009) pp. 1272–1275.

Reijnders, L. and Huijbregts, M.A.J. 'Palm oil and the emission of carbon-based greenhouse gases', *Journal of Cleaner Production,* Vol. 16, (2008) pp. 477–482.

Sefton, M., Shupp, R. and Walker, J.M. 'The effect of rewards and sanctions in provision of public goods', *Economic Inquiry,* Vol. 45, (2007) pp. 671–690.

Sekhon, G. S. 'Fertilizer-N use efficiency and nitrate pollution of groundwater in developing countries', *Journal of Contaminant Hydrology,* Vol. 20, (1995) pp. 167–184.

Sutter, M., Haigner, S. and Kocher, M.G. 'Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations', *Review of Economic Studies,* Vol. 77, (2010) pp. 1540–1566.

Tiedemann, T. and Latacz-Lohmann, U. ‘Production Risk and Technical Efficiency in Organic and Conventional Agriculture - The Case of Arable Farms in Germany’, *Journal of Agricultural Economics,* Vol. 64, (2013) pp. 73–96.

Tute, C. ‘An outcome-based payment scheme for the promotion of biodiversity in the cultural landscape’, *Integrating Efficient Grassland Farming and Biodiversity,* (2005) pp. 36-39.

UNCTAD. ‘Palm oil’, Available at: http://www.unctad.info/en/Infocomm/AACP-Products/Palm-oil/, (2013) (last accessed 21October 2013).

Veetil, V. ‘Conceptions of rationality in law and economics’, *European Journal of Law and Economics,* Vol. 31, (2011) pp. 199–228.

Veldkamp, E. and Keller, M. ‘Nitrogen oxide emissions from a banana plantation in the humid tropics’, *Journal of Geophysical Research: Atmospheres,* Vol. 102, (1997) pp. 15889–15898.

Veldkamp, E., Purbopuspito, J., Corre, M.D., Brumme, R. and Murdiyarso, D. ‘Land use change effects on trace gas fluxes in the forest margins of Central Sulawesi, Indonesia’, *Journal of Geophysical Research: Biogeosciences,* (2008): Vol. 113.

Viceisza, A. ‘Treating the field as a lab: A basic guide to conducting economics’, *Food security in practice technical guide series 7,* International Food Policy Research Institute (IFPRI), (2012).

Volpp, K.G., John, L.K., Troxel, A.B., Norton, L., Fassbender, J. and Loewenstein, G. ‘Financial incentive–based approaches for weight loss’, *JAMA,* Vol. 300, (2008) pp. 2631–2637.

Wilcove, D.S. and Koh, L.P. ‘Addressing the threats to biodiversity from oil-palm agriculture’, *Biodiversity and conservation,* Vol. 19, (2010) pp. 999–1007.

Xe. Available at: http://www.xe.com/currencytables, (2014) exchange rate from 15 August 2012, (last accessed 28 November 2014).

Appendix A: Instructions and agenda

The experiment took place in a public building, e.g., a school, a gymnasium, an administration building or in the house of the village leader. Depending on the size of the village, we invited between 10 and 18 randomly chosen farmers. The experiment started either after lunch or after the evening prayer. When the farmers arrived, they were identified and interviewed. Here, personal as well as economic information from the farmers was gathered. For the experiment, the farmers sat on the floor or, if available, on chairs with tables. To avoid chatting among the farmers, we distanced them from each other. One researcher and three or four enumerators ensured a smooth conduction of the experiment. For the experiment, we used questionnaires, one for each farmer, including a fertiliser-output table (Figure A1), an example sheet (Figure A3) and one decision sheet for each of the ten rounds (Figure A2). Furthermore, to support our explanation, we used posters that were equal to the questionnaire sheets.

The Experiment starts:

Welcome to our experiment and thank you for coming. *We introduced ourselves, our research project and our home university*. This experiment simulates a palm oil plantation being owned and operated by you. The results from this experiment are of academic interest. If you complete it, you will earn a shopping voucher for a local shop. The value of the shopping voucher depends on the decisions you have to make during this experiment, so listen carefully. Whenever you have a question, do not hesitate to ask. The experiment will take approximately one hour and will start immediately.

We distributed the questionnaires to the farmers and hung up posters to support the explanations (see Figure A1, Figure A2 and Figure A3).

We start explaining the simulated situation to the farmers: Imagine you have a palm oil plantation of about one hectare. Your plantation is already established and in production. You have to decide how much fertiliser you want to use, which will affect your output. Then, you will sell this output, pay for the fertiliser and, additionally, you might receive some public money. So, you will end up with a profit. You will receive a shopping voucher in the value of 5% of this profit. To make numbers in the experiment more manageable, we deviate from the realistic prices, incentives and penalties by the factor of 1,000. Now, we will explain to you the experiment in detail.

This experiment simulates a palm oil production for 10 years, where each year equals one round. At the beginning of each year, you have to decide how much fertiliser you want to use. This is the only decision you have to make during the year. You are allowed to use between 0 and 1,000 kg fertiliser, but only in 10 kg increments. The fertiliser used is the only determinant for the output you receive. This means that the experiment does not consider diseases, pests, weather conditions etc. We are aware that this is not realistic, but for the purpose of this experiment, this assumption is necessary. To make the fertiliser decision easier for you, we put a fertiliser-output table in your questionnaire showing the output for a certain amount of fertiliser (Figure A1).

After you have decided how much fertiliser you will use, the price for your output is ascertained. To generate the price for each of the following years, you have to draw a ball from a bag with three green and tree red balls. Starting with a price of 1,800 Rp, the price rises by 200 Rp compared to the previous year if you draw a green ball, whereas the price falls by 200 Rp if you draw a red ball.

When the price is known, an enumerator will calculate your profit for the actual year. Thus, the enumerator declares your output and, based on the price, calculates your revenue. Then the cost for fertiliser, which is 10 Rp per kg, is subtracted. Additionally, there is an unconditional public subsidy of 10,000 Rp per year, which is added. Later on, this public intervention can change. If the profits for all present farmers in a year are calculated, we start with the next year. Again, you decide how much fertiliser you want to use. At the end of the experiment, all profits are added up and you will receive 5% of these profits in the form of a shopping voucher for a local shop.

Take a look at the example provided in your questionnaire (Figure A3). There is a fertiliser use of 100 kg. With a glance at the fertiliser-output table (Figure A1), which is implied in your questionnaire, you can see that this results in a output of 14.5 tons. Furthermore, you can see that the price in the previous round is 1,200 Rp per ton. Since a green ball is drawn in this example, the price increases for 200 Rp to 1,400 Rp per ton. 14.5 tons of output multiplied by a price of 1,400 Rp per ton results in a revenue of 20,300 Rp. To account for the costs of fertiliser, 100 kg multiplied by a price of 10 Rp per kg are deducted. Additionally, there is a subsidy of 10,000 Rp. In total, this results in a profit of 29,300 Rp for this year. 5% of this amount, i.e. 1,465 Rp, would have been added to your shopping voucher.

To check if you understood the experiment, we have some control questions. *We asked the following questions one by one to the audience; if necessary, we repeated our explanation until everyone understood the answer.*

- What kind of plantation do you cultivate in the experiment?
- What is the size of your plantation?
- How many years of palm oil production simulates this experiment?
- If you use 50 kg fertiliser, how much output do you get?
- If you use 150 kg fertiliser, how much output do you get?
- If you want to achieve a output of 20.1 tons, how much fertiliser do you need?
- If the price in the previous year was 1,600 Rp per ton, and you draw a green ball, what is the new price?
- If the price in the previous year was 2,000 Rp per ton, and you draw a red ball, what is the new price?
- What is the cost for 170 kg of fertiliser?
- What amount of subsidy do you receive each year?
- Can the subsidy change during the experiment?

Please ask further questions! If there are no questions left, we can start with the experiment. Please make your own decisions, and do not talk to someone except the enumerators during the experiment. If you talk to other farmers during the experiment, the data is not useful for us and, therefore, you will be excluded from the experiment.

We formed small groups of three to five farmers with one enumerator during the experiment. This eased conduction of the experiment and, if necessary, the communication between the farmers and the enumerator. We conducted round 1-5. We only started a new round if every participant finalised the current round. After the fifth round, the procedure differed between the treatments as follows:

For the control treatment:
Scientists found out that excessive use of fertiliser is bad for the environment. The subsidy of 10,000 Rp per year persists.

For the treatment A:
Scientists found out that excessive use of fertiliser is bad for the environment. Therefore, the subsidy of 10,000 Rp is cancelled and replaced by a new public intervention program which

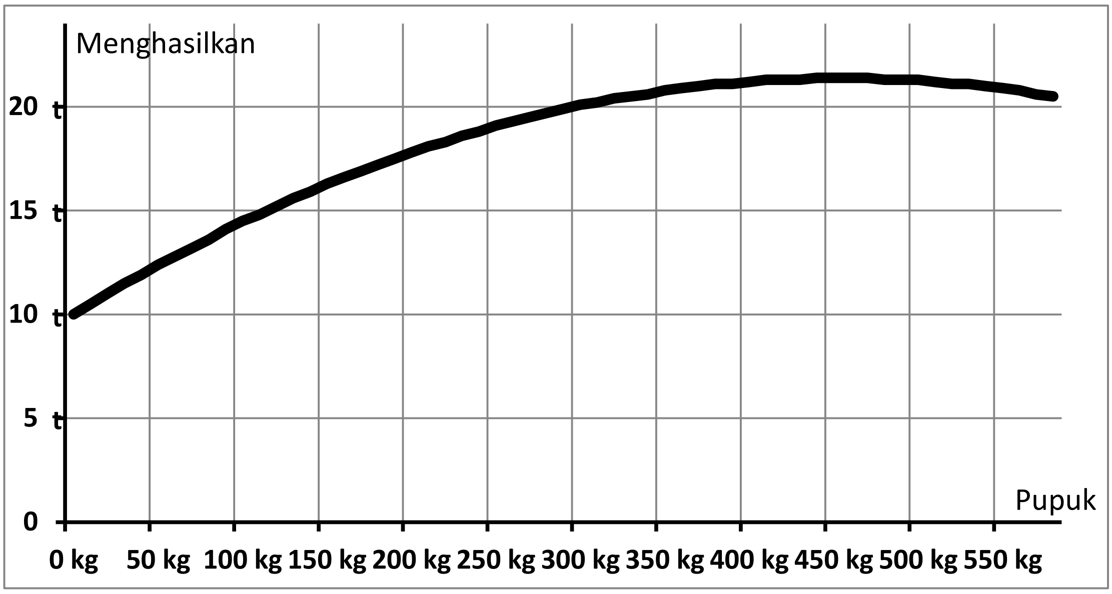
aims to reduce the use of fertiliser. It is a compensation payment of 10,000 Rp for everyone who uses 120 kg or less fertiliser per ha.

For the treatment B:
Scientists found out that excessive use of fertiliser is bad for the environment. Therefore, a new law is introduced aiming to restrict the use of fertiliser. It says that there is a punishment of 10,000 Rp for everyone who uses more than 120 kg fertiliser. The subsidy of 10,000 Rp per year persists.

For the treatment C:
Scientists found out that excessive use of fertiliser is bad for the environment. Therefore, the subsidy of 10,000 Rp is cancelled and replaced by a new public intervention program which aims to reduce the negative externalities generated by the use of fertiliser. There will be a compensation payment of 100,000 Rp for everyone who successfully avoids such negative externalities. If you use 120 kg fertiliser or less, you have a 10% chance to achieve this. To determine if the incentive takes place or not, we use a bag with 9 blue and one yellow ball. If you draw a blue ball, you are not able to avoid the externalities, and you do not receive the compensation payment. If you draw the yellow ball, you are able to avoid these negative externalities, and you receive the compensation payment of 100,000 Rp. The draw is effected at the end of the round after the decision on fertiliser application had been taken.

For the treatment D:
Scientists found out that excessive use of fertiliser is bad for the environment. Therefore, a new law is introduced, aiming to restrict the use of fertiliser. It says that there is a punishment of 100,000 Rp for everyone who uses more than 120 kg fertiliser. The control for this law is not perfect. There is a chance of only 10% to get caught if you use more than 120 kg fertiliser. To determine if you get caught, we use a bag with 9 blue and one yellow ball. If you draw a blue ball, you were not caught, and you do not get any punishment. If you draw the yellow ball, you were caught, and you have to pay a punishment of 100,000 Rp. The draw is effected at the end of the round after the decision on fertiliser application had been taken. The subsidy of 10,000 Rp per year persists.

We conducted round 6 – 10. Subsequently, we summed up the profits and gave the corresponding shopping vouchers to the participants.



Pupuk	Mengha-silkan	Pupuk	Mengha-silkan	Pupuk	Mengha-silkan	Pupuk	Mengha-silkan
0 kg	10,0 t	150 kg	16,3 t	300 kg	20,1 t	450 kg	21,4 t
10 kg	10,5 t	160 kg	16,6 t	310 kg	20,2 t	460 kg	21,4 t
20 kg	11,0 t	170 kg	16,9 t	320 kg	20,4 t	470 kg	21,4 t
30 kg	11,5 t	180 kg	17,2 t	330 kg	20,5 t	480 kg	21,3 t
40 kg	11,9 t	190 kg	17,5 t	340 kg	20,6 t	490 kg	21,3 t
50 kg	12,4 t	200 kg	17,8 t	350 kg	20,8 t	500 kg	21,3 t
60 kg	12,8 t	210 kg	18,1 t	360 kg	20,9 t	510 kg	21,2 t
70 kg	13,2 t	220 kg	18,3 t	370 kg	21,0 t	520 kg	21,1 t
80 kg	13,6 t	230 kg	18,6 t	380 kg	21,1 t	530 kg	21,1 t
90 kg	14,1 t	240 kg	18,8 t	390 kg	21,1 t	540 kg	21,0 t
100 kg	14,5 t	250 kg	19,1 t	400 kg	21,2 t	550 kg	20,9 t
110 kg	14,8 t	260 kg	19,3 t	410 kg	21,3 t	560 kg	20,8 t
120 kg	15,2 t	270 kg	19,5 t	420 kg	21,3 t	570 kg	20,6 t
130 kg	15,6 t	280 kg	19,7 t	430 kg	21,3 t	580 kg	20,5 t
140 kg	15,9 t	290 kg	19,9 t	440 kg	21,4 t	590 kg	20,4 t

Figure A1. Fertiliser-output table; Translation: menghasilkan = production of palm oil bunches, pupuk = fertiliser
Source: Author's own illustration according to Indonesian standard

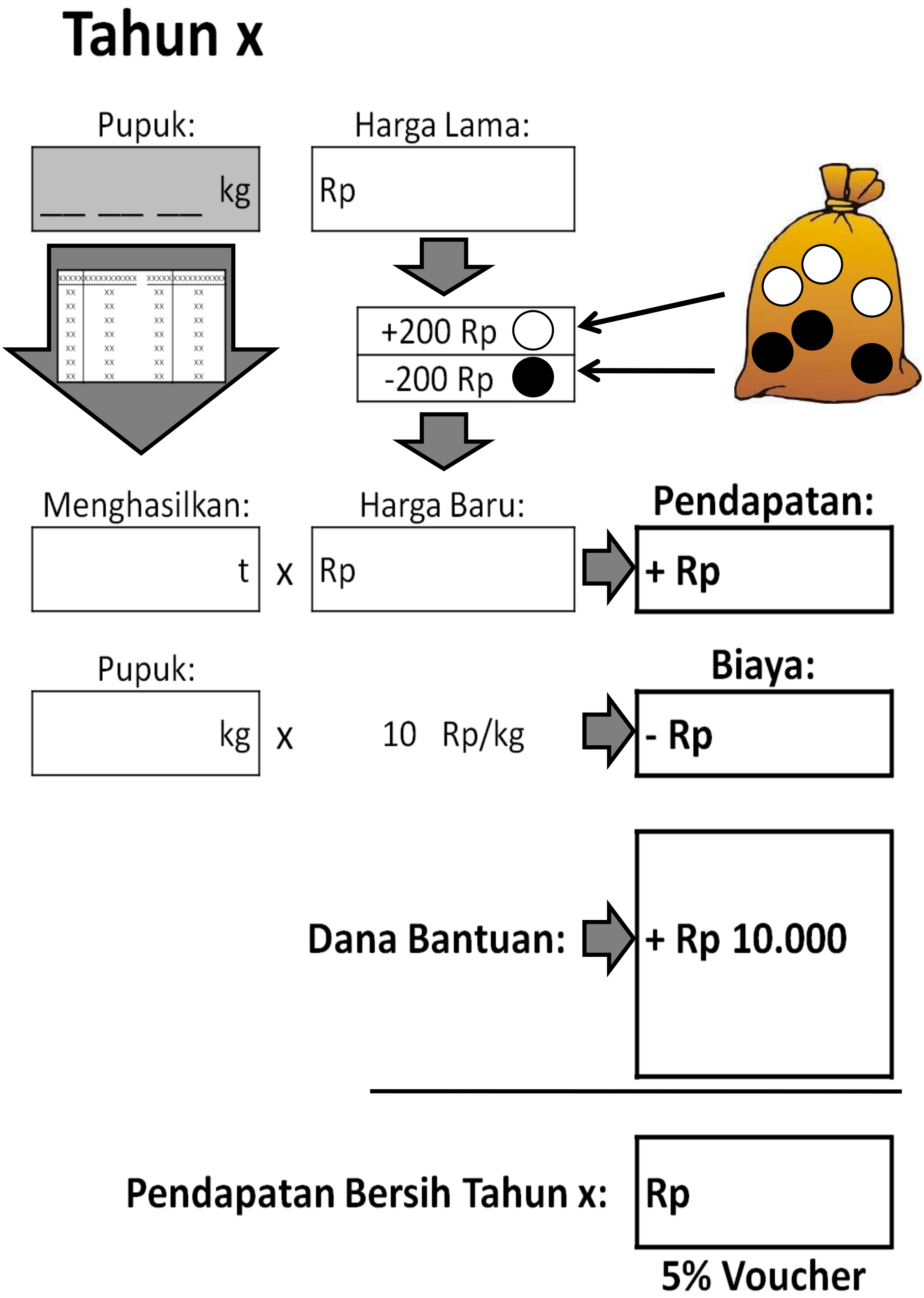


Figure A2. Decision sheet; Translation: tahun x = year x, pupuk = fertiliser, harga lama = old price, menghasilkan = production of palm oil bunches, harga baru = new price, pendapatan = revenue, biaya = cost, dana bantuan = relief fund, pendapatan bersih tahun x = net income year x
Source: Author's own illustration according to Indonesian standard

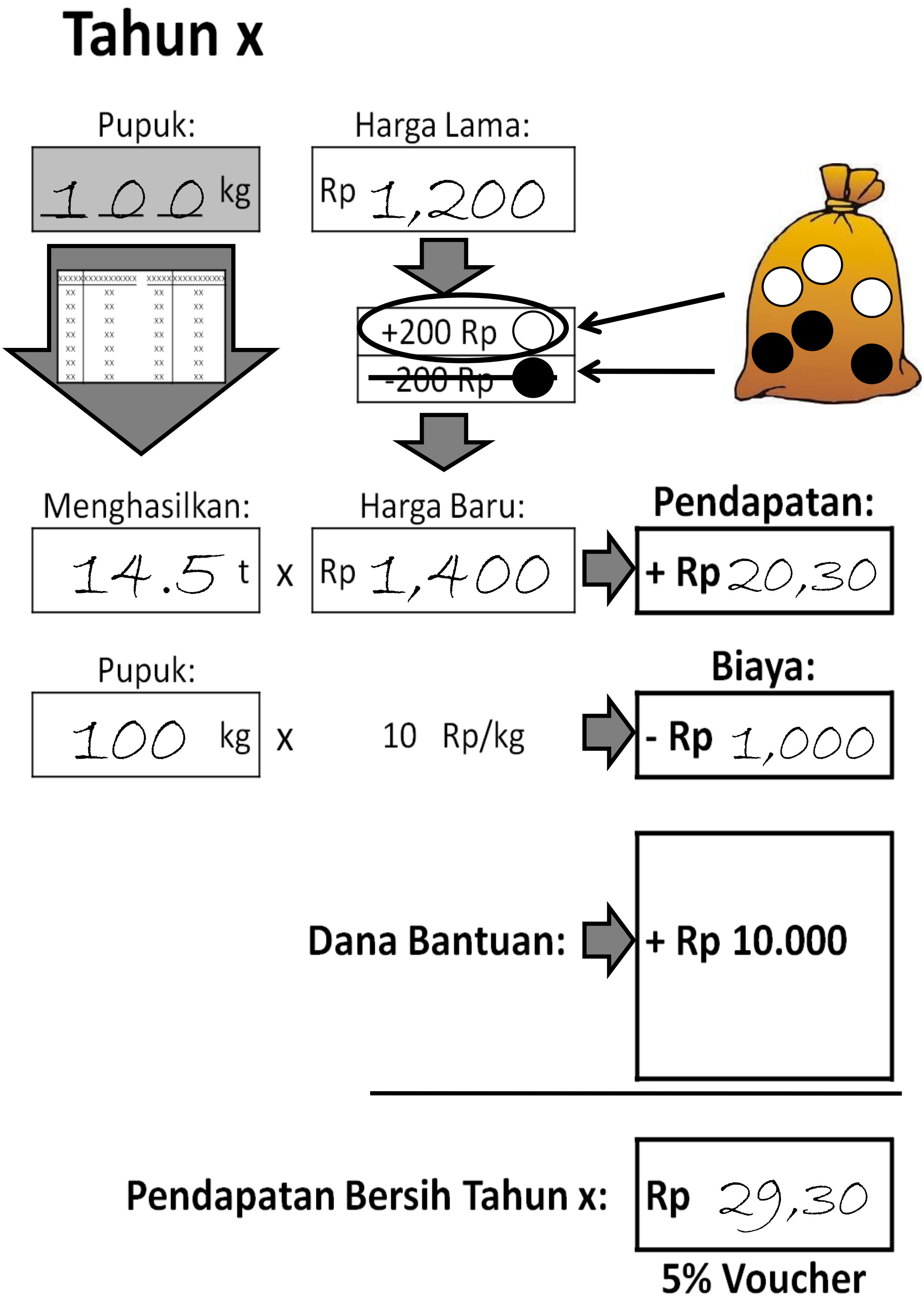


Figure A3. Decision sheet: example; Translation: tahun x = year x, pupuk = fertiliser, harga lama = old price, menghasilkan = production of palm oil bunches, harga baru = new price, pendapatan = revenue, biaya = cost, dana bantuan = relief fund, pendapatan bersih tahun x = net income year x
Source Author’s own illustration according to Indonesian standard

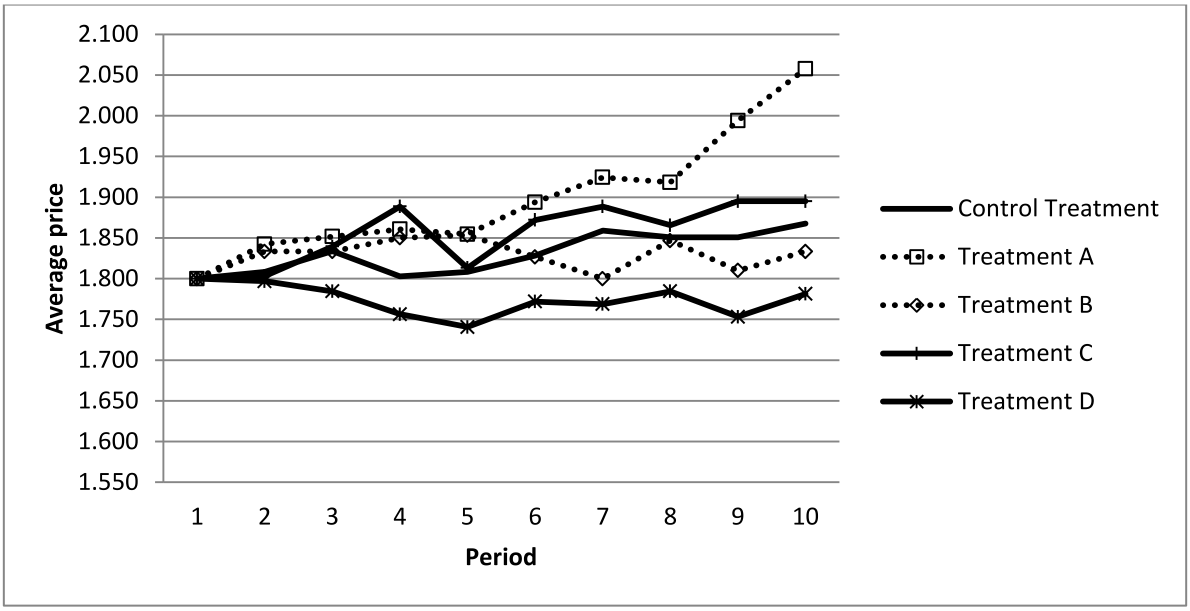


Figure A4. Average prices per treatment
Source: Author's own illustration

Table A1.
Profit-maximising fertiliser use for rational participant

Price	Profit-maximising fertiliser round 1-5, control treatment	Profit-maximising fertiliser round 6-10, treatment A, B, C, D
0	0	0
200	10; 20; 30	10; 20; 30
400	210; 230; 250	120
600	300	120
800	350	120
1,000	350; 360; 370; 380	120
1,200	380	120
1,400	380	120
1,600	410	120
1,800	410	120
2,000	410	120
2,200	410	410
2,400	410	410
2,600	410	410
2,800	410	410
3,000	410; 440	410; 440
3,200	440	440
3,400	440	440
3,600	440	440

Note: It is possible that more than one value result in the profit maximising use of fertiliser