European Commission

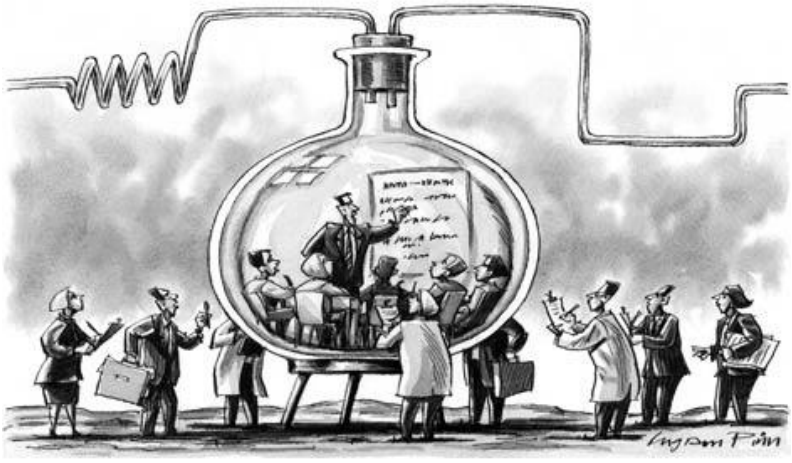
JRC SCIENCE AND POLICY REPORT

(How) can economic experiments inform EU agricultural policy?

Liesbeth Colen[1]
Sergio Gomez y Paloma[1]
Uwe Latacz-Lohmann[2]
Marianne Lefebvre[3]
Raphaële Préget[4]
Sophie Thoyer[5]

[1]European Commission, JRC-IPTS, Seville, Spain
[2]Kiel University, Germany and University of Western Australia, Australia
[3]Université d'Angers, GRANEM UMR, France
[4]INRA, UMR LAMETA, France
[5]Montpellier SupAgro, UMR LAMETA, France

2015



Report EUR 27496 EN

Joint Research Centre

European Commission

Joint Research Centre

Institute for Prospective Technological Studies

Contact information

Liesbeth Colen

Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: jrc-ipts-secretariat@ec.europa.eu
Tel.: +34 954488318
Fax: +34 954488300

https://ec.europa.eu/jrc
https://ec.europa.eu/jrc/en/institutes/ipts

JRC97340

EUR 27496 EN

ISBN 978-92-79-52251-2 (PDF)

ISSN 1831-9424 (online)

doi:10.2791/17634

Luxembourg: Publications Office of the European Union, 2015

Abstract

This report provides an overview of the potential contribution of economic experiments to the Common Agricultural Policy evaluation toolbox. The methodology of economic experiments is briefly described in chapter 2. Chapter 3 presents examples of experimental studies relevant for the analysis of the Common Agricultural Policy. Chapter 4 discusses whether there is room for economic experiments in the EU evaluation toolbox. Chapter 5 identifies the main challenges when implementing economic experiments. Chapter 6 provides a set of recommendations, arising from the discussions between the authors of this report and policy officers from DG AGRI, relative to the introduction of economic experiments in the CAP evaluation toolbox.

Acknowledgements

The authors would like to thank the participants of the workshop "*How can economic experiments inform EU agricultural policy?*", held on 8 January 2015 at DG Agriculture and Rural Development (DG AGRI) of the European Commission, Brussels, for the fruitful discussion and useful suggestions (see Annex 2, *infra*). Emanuele Ciriolo, Koen Dillen, Patrizia Ganci, Annette Hurrelmann, Andreas Kolodziejak, Yves Plees and Rene van Bavel are acknowledged for their useful comments on the report.

Disclaimer

Content

Abbreviation list

AES: Agri-Environmental Scheme

CAP: Common Agricultural Policy

DCE: Discrete Choice Experiment

DG: Directorate-General

DG AGRI: Directorate-General for Agriculture and Rural Development

EC: European Commission

EU: European Union

FADN: Farm Accountancy Data Network

RCT: Randomised Control Trial

RDP: Rural Development Policy

WTP: willingness-to-pay

WTA: willingness-to-accept

PSM: propensity score matching

DID: difference-in-difference

Executive summary

The last two decades have seen a rapidly growing interest in the use of experiments to analyse, test and evaluate policies, both at national and European level. In the European Commission (EC), several DGs have already commissioned experimental studies in the fields of consumption, health, employment and the environment. Compared to these other fields, the use of experimental methods for evaluating the Common Agricultural Policy (CAP) has remained very limited up to now.

This report provides an overview of the potential contribution of economic experiments to Common Agricultural Policy evaluation.

Along the report, "evaluation" is used in a broad sense, including both ex-ante (impact assessment in EC terminology) and ex-post evaluation (evaluation in EC terminology).

This report follows up on the presentations and discussions at the workshop "*How can economic experiments inform EU agricultural policy?*", organised by the Joint Research Centre and held on 8 January 2015 at DG AGRI of the European Commission, Brussels.

What are economic experiments?

Experiments are procedures where economic questions are studied by making participants go through a particular exercise or "game", designed to capture the decision situation under study. The data collected in such experiments can for example be used to test economic theories, to reveal economic decision making in the presence of policies, and to reveal market mechanisms. Experiments can be conducted in the laboratory, in the field or online, but the experimental data generation process always involves two main features:

First, participants in an experiment are randomly allocated to a treatment group, designed to represent the policy under evaluation, or a control group, representing the current policy or status-quo.

Second, participants are usually incentivised to reveal their true preferences: they receive a payment according to their decisions in the experiment.

These general characteristics hide a large variety in economic experiments. In the report, we classify experimental approaches in four categories: discrete choice experiments (DCE), laboratory experiments, field experiments and randomised control trials (RCTs).

Is there room for economic experiments in the EU agricultural policy evaluation toolbox?

The Common Agricultural Policy (CAP) has experienced several reforms over the last decades. In order to meet the CAP's evolving objectives, the type of policy interventions has changed drastically. As a result, also the toolbox for evaluating these policies needs to be constantly adapted to follow these policy changes.

The CAP toolbox for ex-ante policy assessment and ex-post evaluation studies includes: farm and market simulation models, calibrated with EU-wide statistical data (e.g. FADN or FSS), statistical and econometric analysis of survey data and a range of qualitative methods such as interviews with stakeholders, focus group or internet-based public consultation (EC 2015).

Recent evolutions in the CAP call for evaluation methods able to: (1) capture the change in support target, from commodities to farm holding, therefore using the farm holding as the main unit of policy evaluation; (2) account for the voluntary aspects of many measures, (3) allow the evaluation of purely regulatory measures (with no EU funding involved) and (4) flexible enough to account for the heterogeneity in CAP implementation due to the greater flexibility given to Member States for implementing first and second-pillar measures.

In this context, economic experiments can be a useful addition to the CAP evaluation toolbox in three directions:

First, experiments are particularly well suited to shed light on the behavioural drivers of farmers' responses to policy. In that direction, experimental data can be combined with non-experimental evaluation tools. For example, experiments allow the elicitation of behavioural parameters (e.g. risk aversion, time preferences) which can then be included in simulation models. Also, they can reveal information to test whether other types of objectives functions would reflect better farmers' preferences.

Second, experimental approaches allow testing different variations in policies prior to implementation in order to compare their effectiveness and/or efficiency. Especially when moving to new types of policies, simulation or econometric models based on data observed in the past will do a poor job in predicting farmers' reactions to a new policy. When it comes to targeted policy interventions (on specific farm types or in a limited number of regions), simulation models are not the most appropriate method either, given the cost of making the models flexible enough to capture policy specificities at the regional or farm level and because of the lack of data to feed such models.

Third, economic experiments, especially RCTs, can contribute to ex-post policy evaluation, by providing clean information on the counterfactual situation. Thanks to the high control in the data collection process, experimental approaches allow to measure the net impact of the policy and to distinguish it from other factors, which is not always possible with methods based on observational data. But RCTs provide limited insights on the reasons behind the observed effects.

Following the Introduction (chapter 1) and an overview of economic experiments for policy evaluation (chapter 2), chapter 3 reviews the most noteworthy applications of economic experiments already available in the literature and likely to inform EU agricultural policy-makers. This chapter provides an illustration of the possibilities of the different types of experimental approaches to evaluate agricultural policies.

Complementarities between experimental and non-experimental methods are further discussed in chapter 4. More precisely, experimental approaches are compared to other evaluation methods in the light of the trade-off between "internal validity" (i.e. the ability to identify the causal relation between two variables, for example the policy and the outcome) and "external validity" (i.e. the ability to transfer the results to the real world and to other contexts). In lab experiments, internal validity is guaranteed by the high level of control of the decision-making environment and random assignment of participants to the different treatments, which allows preventing systematic differences in treatment and control groups. However, the ability to transfer the results to the real world could be an important limitation of lab experiments. Field experiments, RCTs

and DCEs present a compromise and can lessen the inherent tension between establishing causality and guaranteeing the generalizability of results.

The different methods are also compared with respect to their practical implementation, such as the ease of data collection, and the ease of interpretation of evaluation results by policy makers.

Main practical issues associated with the setting-up of experiments

The first difficulty concerns the recruitment of participants. The fact that many types of experiments cannot be organised via the internet or at the respondents' home creates extra difficulties compared to traditional surveys. As a result, sample size may be limited and/or the risk of selection bias increased, and this will often call for compromise in the search for sample "representativeness".

Second, when the experimental approach requires setting up sub-groups receiving different treatments, particular attention must be paid to the procedure assigning participants to treatment and control groups.

Third, randomisation can raise ethical concerns since it may create inequities across participants. "Close to random" procedures are described. They can be used to introduce some randomisation elements in experiments (thereby reducing selection bias) while increasing acceptance compared to pure randomisation.

While experiments generally provide more robust evidence than traditional surveys or stakeholder consultation, strategic biases can still be observed. Such strategic bias arises when participants anticipate that their responses or observed behaviour will have an impact on future policy choices and may want to manipulate the outcomes of the experiment by adopting an insincere behaviour. Such bias must be controlled by a careful design of the experimental protocol.

Finally, identification and recruitment of participants impose organisational constraints, which in turn induce costs. The costs depend on the scope of the study and the type of experimental design. As a general rule of thumb, laboratory experiments can be expected to be cheaper and yield more rapid results than field experiments and DCEs, whereas RCTs often involve the most costly and time-consuming set-up.

The incorporation of experimental approaches in the current CAP evaluation toolbox may not be straightforward. A list of six recommendations, that have benefitted of comments by selected DG AGRI policy officers (see Acknowledgments), is provided to enhance the use of experimental approaches in the CAP evaluation process.

Recommendations to evaluators when proposing the use of experiments for CAP analysis

1. Make use of the complementarity of methodologies and be open to the use of innovative methods

Policy officers are fully aware of the need for the toolbox to evolve. Take advantage of the absence of constraints in the methodology in most calls for tenders to propose innovative methods, including experimental approaches. The complementarities between economic experiment and non-experimental approaches are numerous and should be highlighted.

2. Reconsider the search for "representativeness"

Policy officers generally look for evidence general enough to be valid for the entire population under study. However, given the limitations in sample size and the sampling issues involved in experimental studies, this will not always be possible. The representativeness and generalizability of the results should be discussed in detail to avoid misunderstanding. Repeating experiments in a number of different contexts (different regions, farm types) can also increase the robustness and acceptability of experimental results by policy officers.

3. Proceed by incremental steps: from the laboratory to the field.

Different types of experiments can complement each other. For example, new policies could be first tested in the lab and then in the field, before moving to randomised control trials.

4. Make sure experiments are implemented in a non-discriminatory way.

It is crucial that any EU action is implemented in a non-discriminatory way. Also when making use of experiments, one should make sure to respect this principle. Innovative approaches to randomisation and close-to-random procedures may provide a solution.

Recommendations to EC policy officers to enhance the use of experiments

1. Invest in resources and expertise on the use of experimental methods for policy evaluation

With respect to the development of policy proposals and their ex-ante impact evaluation, investing in internal (EC) expertise on the use of experimental methods would allow spreading their use to inform policy making. One example is the newly created 'Unit for Scientific Foresight and Behavioural Insight' of the Joint Research Centre of the EC.

DG AGRI can also make use of the *Framework Contract for the Provision of Behavioural Studies* to commission experiments on the behavioural aspects of farmers' policy responses.

2. Inform evaluators on the potential of experimental approaches and openness towards the use of these new methods

Informing external contractors on the advantages and limitations of experimental approaches for policy making could help to show DG AGRI's openness towards the use of these methods in future evaluation contracts. Also Member States could be informed on the potential of experimental approaches, for example with respect to the evaluation of nationally implemented rural development policies.

Moreover, informing evaluators on the procedure for authorisation of pilot tests and randomisation by the EC services would allow evaluators to take into account the constraints associated to these procedures and may encourage the setting-up of RCTs.

1 Introduction

European policies are constantly analysed, monitored and evaluated, allowing their adjustment or reform on the basis of sound evidence. The need for policy evaluation arises both before and after policy implementation (Figure 1).

The methodological toolbox for policy evaluation has made several important advances in the last decades. First, new tools for identifying the causal impact of policies have been developed, including both experimental and quasi-experimental approaches (EC 2013a). Second, behavioural studies have shed light on the need to account for many elements of the decision context, beyond simple profit maximisation, in order to predict economic agents' responses to different policy instruments.

In this report, we investigate how such advances can be further integrated into the evaluation toolbox of the European Commission (EC). More precisely, we focus on the relevance of experimental methods for the evaluation of the Common Agricultural Policy (CAP) and their potential contribution to evidence-based agricultural policy making (see box 1 for a clarification on terminology).

In agriculture, experiments have been used for a long time. Field trials are widely used by agronomists to measure differences in agricultural productivity of different practices (Splawa-Neyman 1923 [traduced in 1990]; Fisher 1926). Economic experiments attempt to apply the same principles to human participants and their economic decisions. But the usefulness of experiments in economic research and policy analysis was initially thought to be limited, because of the difficulty of controlling the wide range of factors that influence human decision making.

> « Economists cannot perform the controlled experiments of chemists or biologists because [they] cannot easily control other important factors. Like astronomers or meteorologists, [economists] generally must be content largely to observe » (Samuelson and Nordhaus 1985).

Yet, experimental methods have been further developed to observe and analyse the behaviour of economic agents under controlled institutional environments and to inform policy makers about the impacts of current and proposed policy alternatives (Kagel and Roth 1995). As a result, the last two decades have seen a rapidly growing interest in the use of experiments to test and evaluate policies, both at national and EU level.

Box 1: Terminology used in the report

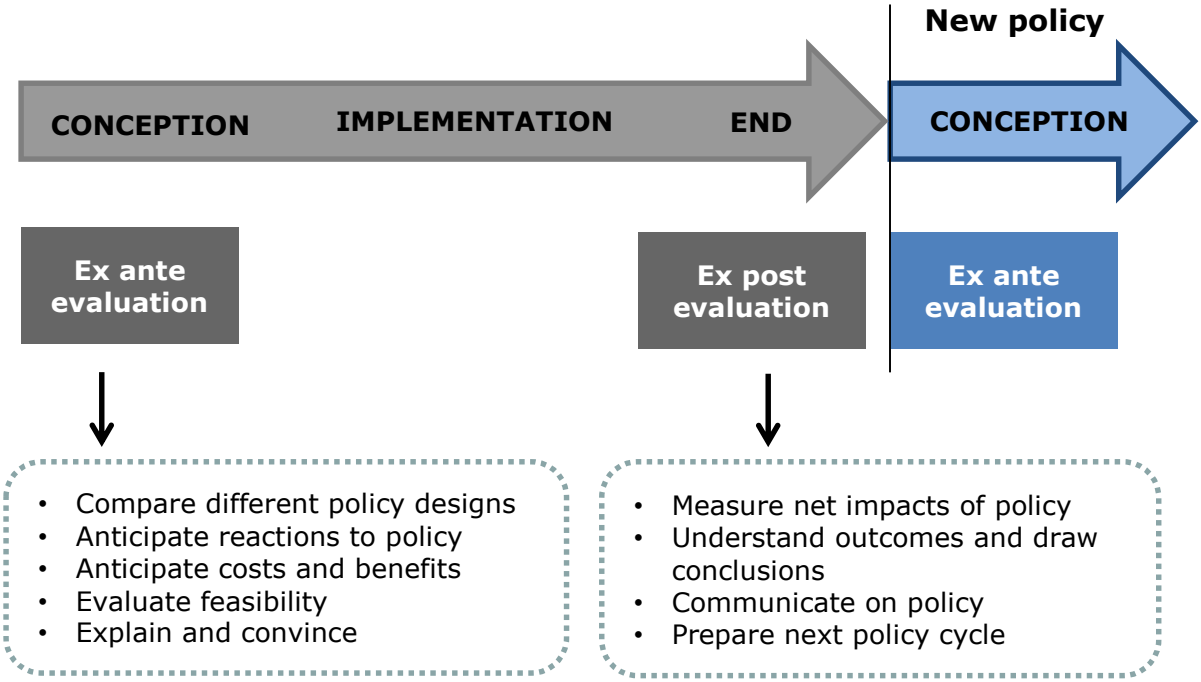
Throughout this report, "evaluation" is used in a broad sense, including both ex-ante (impact assessment in EC terminology) and ex-post evaluation (evaluation in EC terminology).

Ex ante evaluation contributes to the choice of the most efficient or cost-effective policies among different options. For example, the European Commission performs an Impact Assessment for all initiatives that are likely to have significant economic, environmental and social impacts, consisting of a careful analysis of the potential impacts of different policy scenarios before a policy proposal is made (EC 2015).

Ex post, it is important to know if the policy has reached the expected targets or not, and to identify the reasons for success or failure. In the EU, the CAP is evaluated regularly through evaluations targeted at specific measures and based on the implementation of the policy over the last years.

While this distinction between ex ante and ex post is useful for evaluators and researchers who need to select the most appropriate methodology, this distinction may be less straightforward in practice given that time is continuous and policy implemented at time t also serves as input for ex ante evaluation of the next policy at time t+1.

Figure 1: The policy evaluation cycle (Source: Authors' own elaboration)



In the European Commission, several Directorates-General (DGs) have already commissioned experimental studies to analyse the impact of policies in the fields of consumption, health, employment or the environment (Ciriolo 2011; van Bavel, Rodríguez-Priego, and Maghiros 2015). Compared to these other fields, the use of

experimental methods for evaluating the CAP has remained very limited up to now.[1]

Could the experimental approach be therefore a relevant addition to the CAP evaluation toolbox?

Experimental methods are useful only if they can complement other evaluation methodologies. There are already many powerful methods for agricultural policy evaluation in Europe, such as the analysis of high-quality agricultural statistics (FADN, Eurostat Farm Structure Survey), qualitative case-study approaches, and several economic models representing the farming sector, both at market and farm levels. Therefore, the focus of experimental studies should be on how they can add robust evidence to the existing impact assessment and evaluation process (Rodríguez-Priego and van Bavel 2014).

This report follows up on the presentations and discussions at the workshop "*How can economic experiments inform EU agricultural policy?*", held on 8 January 2015, organised by the Joint Research Centre and housed by DG Agriculture and Rural Development of the European Commission, Brussels. The objective of the workshop was twofold: i) to introduce DG AGRI staff to economic experiments as a tool for agricultural policy design and evaluation; ii) to discuss how economic experiments can complement existing CAP evaluation methods. The agenda for the meeting and list of participants are presented in Annex 1.

The methodology of economic experiments is briefly described in chapter 2. Chapter 3 presents examples of experimental studies relevant for the analysis of the Common Agricultural Policy. Chapter 4 discusses whether there is room for economic experiments in the EU evaluation toolbox. Chapter 5 identifies the main challenges when implementing economic experiments. Chapter 6 provides a set of recommendations, arising from the discussions between the authors of this report and policy officers from DG AGRI, relative to the introduction of economic experiments in the CAP evaluation toolbox.

[1] In an article entitled "Can field experiments return agricultural economics to the glory days?", Herberich, et al. (2009) speculate that the future of field experimental methods can be importantly engineered by agricultural and resource economists.

2 Economic experiments for policy evaluation: an overview

2.1 Definition and classification of experimental approaches

Economic experiments are procedures whereby participants are made to go through a particular exercise or "game", designed to capture the situation under study (van Bavel, Rodríguez-Priego, and Maghiros 2015). The data collected in such experiments can for example be used to test economic theories, to reveal economic decision making in the presence of policies, and to reveal market mechanisms. Experiments are either conducted in the laboratory, in the field or online.

Experiments consist in data generation rather than the collection of observational (or "naturally occurring") data (Herberich et al. 2009). This data generation process is controlled by the analyst and often includes two main features:

First, participants in an experiment are in most cases randomly allocated to one of the treatment groups, designed to represent the policies or instruments under evaluation, or a control group, representing the current policy or status-quo. Their behaviour is then observed and compared to that of participants exposed to different interventions (or to nothing, i.e. the control group).

Second, participants are affected by the decisions they make in the experiment: they usually receive a payment according to their decisions in the experiment. The incentives are not just for showing up. Rather, they guarantee that participants take their tasks seriously and thereby emulate real-life behaviour as much as possible (van Bavel, Rodríguez-Priego, and Maghiros 2015). Ideally, the magnitude of the financial incentives in the experiment should be close to those in the real world to ensure that participants' decisions in the experiment correspond to those in real life.[2] Decisions are therefore less biased by strategic responses compared to answers to hypothetical questions. An experimental design is said to be incentive-compatible if all of the participants fare best when they truthfully reveal their preferences. When such mechanisms are used, experiments are classified as revealed preferences approaches.[3]

These general characteristics hide a large variety in economic experiments described below.

2.1.1 Discrete choice experiments

DCEs are used for assessing people's preferences and/or decisions in hypothetical situations, e.g. before a new product is launched, a new technology becomes available, or a new policy is implemented.

[2] One limit of the experimental approaches is the difficulty to replicate 'real losses'. In very few experiments participants can incur losses beyond the dotation which may be given in the experiment.

[3] Revealed preferences oppose to stated preferences, where agents are asked to declare their preferences in a hypothetical setting.

DCE are designed such that it is possible for the respondent to envisage the different situations in the real world even when they do not exist yet.

A DCE consists of several rounds. In each round, respondents are presented with a 'choice set' of alternatives, and are asked to indicate their preferred option. Typically each alternative consists of a combination of characteristics, called 'attributes', one of which is a price (cost) or a payment (benefit). Hence, when making their choice, respondents have to trade-off between the lower monetary benefit or higher cost of one alternative, and the more preferable non-monetary characteristics of another alternative. The results of these repeated choices allow estimating in monetary terms the value of each attribute for the participants. Or, more specifically, what their willingness-to-pay (WTP) or willingness-to-accept (WTA) is for changes in the level of each attribute.

A DCE can thus provide estimates of welfare gains or losses: for instance, it can provide an estimate of the value the public places on an improvement or deterioration in a specific element of their environment, or the costs incurred by farmers for adopting a set of environmentally friendly land management practices. Well-designed DCE usually involve context-specific lists of attributes and values for these attributes. As a result, these estimates can usually not be generalised to other contexts or populations.

Unlike all other experimental methods presented here, DCE is usually a stated-preference method[4]: respondents are asked to choose amongst hypothetical scenarios and do not receive incentive payments for their responses. Thus, it may be that these 'stated' preferences do not correspond with their 'revealed preferences', i.e. the choices they would effectively make in real life.

Nevertheless, DCEs present several advantages over alternative formats of contingent evaluation methods such as open-ended questions (which also provide only stated preferences). Indeed, DCE are a more sophisticated method which reduces bias in responses by making it more difficult for the respondent to respond strategically.

2.1.2 Lab experiments

Lab experiments take place in a laboratory, often a computer room, where participants are invited to take part in an exercise or game which aims to test economic behaviour in response to a set of different scenarios. [5] Participants are financially incentivised to mimic real-world incentives: they can win or lose money according to their decisions in the experiment and are paid in accordance to their choices or performance at the end of the experiment. Participants are confronted with choices that are usually formulated in an abstract way. Such a decontextualized setting allows the experimenter to ensure that the results of the experiment are not influenced by the way the game is framed.

[4] Real discrete choice experiments, where discrete choice questions are combined with real economic incentives, are developing fast to analyze consumers' preferences and willingness to pay for new products (Lusk and Schroeder, 2004; Alfnes et al., 2006; Michaud et al., 2012; Ginon et al., 2014). However, they are more difficult to set-up when the choice under study is a policy option rather than a consumption good.

[5] A video presenting lab experiments produced for the workshop is available at https://www.youtube.com/watch?v=XuJ-cu32z3k

Laboratory experiments are used to pre-test new instruments or policy designs. Different treatments of a lab experiment allow comparing several policy options: for example, how behaviour changes when the implementation rules of a policy are changed, or when participants are allowed to communicate or cooperate before taking their decision, and whether providing different information sets affects participants' decisions.

Thanks to randomisation of participants to alternative treatments, and the high control over the decision context in the lab, lab experiments are valued for their almost clinical ability to isolate interventions and measure their actual impact on behaviour, beyond simply establishing correlations (van Bavel, Rodríguez-Priego, and Maghiros 2015).

Most lab experiments are conducted with students. When real stakeholders from the field are recruited as participants, this is referred to as an *artefactual field experiment* in the literature (Harrison and List 2004).

2.1.3 Field experiments

While lab experiments have been used more widely - partly because they are easy to implement, not very expensive, and students are easy to recruit - field experiments have recently gained popularity in the domain of (agricultural) economics (Herberich et al. 2009).

While lab experiments allow for a perfect control of the decision environment, the artificial context of lab experiments may result in behaviour that is poorly correlated with naturally occurring behaviour. As such, for analysing certain topics, lab experiments may be of little relevance for predicting behaviour in the field (Harrison and List 2004). When the context is believed to impact on participants' decision making, one may prefer to move towards a more realistic 'field' setting, where the experiment is designed such as to represent as closely as possible the field context and the policy to be assessed. This is referred to in the literature as a *framed field experiment.*[6] In such experiments the commodity used, the terminology, the task to be executed or information set that the participants use, correspond to the real environment in which the task or policy takes place.

2.1.4 Randomized Controlled Trials

A last class of experiments are the Randomised Controlled Trials (RCT), sometimes also referred to as 'Randomised Field Trials' (Burtless 1995). The RCT methodology has been developed and progressed in clinical research. Its diffusion into the social sciences and, in particular, economics has been slow, but in recent years economic RCTs have become more frequently used, especially in the fields of labour and development economics.

An RCT is a study design that randomly assigns participants to a treatment group or

[6] Following the classification proposed by Harrison and List (2004). They also include *natural field experiments* in their classification of field experiments, which they refer to as "a framed field experiment that takes place in the environment where the participants naturally undertake these tasks and where the participants do not know they are taking part in an experiment". In this report, our field experiments category only refers to experiments in which subjects play 'games' (with or without interactions with other subjects) and get paid according to their decisions but their participation does not have any consequences in their real life. Hence, they always know they are taking part in an experiment. We instead consider *natural field experiments* to be closer to randomized controlled trials, which are discussed as a separate category (2.1.4).

a control group, where the treatment consists of the real implementation of a policy or programme. Hence, participants are not asked to act as if they would face a certain policy, or are not asked to play a game with financial incentives. Instead, an RCT actually implements the policy among a group of randomly selected 'true' stakeholders. It then observes the actual behaviour of the participants under this policy and compares it to a group of comparable participants that did not receive the policy treatment.

Random allocation means that all participants have the same chance of being assigned to either the treatment or the control group. The purpose of random allocation of participants is to assure that, before the start of the intervention (also called the baseline), participants in both groups have the same characteristics. As such, any observed group differences at the end of the study cannot be due to differences in the average person in each group, for the assignment process renders such differences unlikely. Random assignment provides the best counterfactual describing what would have happened to treatment group participants had they not been exposed to the treatment (Cook, 2003).

For a more comprehensive description of the RCT methodology, the reader is referred to Duflo, Glennester & Kremer (2007).

Table 1: Classification of economic experiments

	Participants Nature of the participants' pool	**Context** Nature of the task or institutional rules applied, and the environment that the participants operate in.	**Incentive mechanism** The mechanism used to allow the revelation of preferences.
Discrete Choice experiments	Participants drawn from the population of interest (e.g. farmers)	Respondents are presented with different choice cards and, in each choice card, they have to select their preferred option among different contextualised scenarios or alternatives	No incentive mechanism in hypothetical DCE. In non-hypothetical DCE, participants bear the financial consequences of their choice.
Laboratory experiments	Students	The context is artificial or synthetic in certain dimensions (decontextualised experiment).	Payments according to choices or performance in the experiment
Field experiment	Participants drawn from the population of interest (e.g. farmers)	In artefactual field experiments, the context is artificial or synthetic in certain dimensions, as in lab experiments. Framed field experiments are framed in the field context of the commodity, task, stakes, or information set of the participants (contextualised experiment).	Payments according to choices or performance in the experiment
Randomised Control Trial	Participants drawn from the population of interest (e.g. farmers)	Participants are randomly assigned to the control or treatment group. For the control group, the environment is the one in which participants are naturally undertaking certain tasks. For the treatment group, one element of the natural environment (for example a policy rule) is changed.	Participants are usually not paid for their participation, but get the benefit from the treatment.

Source: Authors' own elaboration, inspired by Carpenter et al. (2004) and Harrison and List (2004)

2.2 Main uses of experiments

We review here the three main uses of experimental approaches: (1) measuring the causal net impact of policies; (2) highlighting the importance of behavioural factors, and (3) testing policies before implementation.

2.2.1 Generate fine-tuned data enabling measuring the causal net impact of the policy

"Controlled" experiments resolve the major identification issue of empirical analysis based on field data: demonstrate causality and eliminate selection bias.

Ability to demonstrate causality

The standard challenge of policy evaluation is identifying the causal impact of the policy, i.e. the difference between an outcome with the policy (data observed) and the outcome that would have been observed in the absence of the policy (counterfactual).

For example, after implementation of a policy to support farm investments, one may observe that farmers have obtained higher yields. Yet, if over the same period the price of fertilizer has dropped, part of the yield increase may be due to increased fertilizer use. To understand which part of the yield increase can be attributed to the policy, one needs to consider what would have happened in a situation without the policy. To do so, information would ideally be available on beneficiaries of the policy and on those same individuals in a situation without the policy.

Experiments allow creating such a set-up where one group (the treatment group) is exposed to the policy, while the other is not (the control group). As such, any differences in the results indicator between the two groups can be attributed to the causal effect of the policy. Thus, by strictly controlling who is and who is not affected by the policy, experiments can isolate the causal policy impacts from any other factors that may change at the same time.

Ability to eliminate possible selection bias

A second issue is related to the self-selection of beneficiaries. Self-selection implies that only certain types of farmers or regions may choose to participate in a given programme or are programme-eligible. The challenge of evaluation is to distinguish the impact of the policy from the other pre-existing differences between programme participants and non-participants. For example, suppose the farmers soliciting investment support in a region are those farmers with larger gross margin per hectare. A comparison between supported farmers and the others is likely to show a higher gross margin for the supported farmers. But this would not only be thanks to support, but also due to the fact that only the most profitable farms are the ones having received investment support.

Experiments allow eliminating such selection bias by randomising participants across treatments: each person in the group of potential beneficiaries is randomly assigned to either the group that will be exposed to the policy or to the group that does not receive the policy. Because of this random assignment, the participants in both the beneficiary and the 'control' group should have similar (observed and unobserved) characteristics.

By addressing causality and selection bias, experimental data can produce an unbiased

measure of the causal net effect of the policy[7].

2.2.2 Elicit economic agents' preferences and understand their reactions to policy in the presence of behavioural factors

Behaviours are guided by perceptions, impulses and emotions; display biases and heuristics; evidence limited cognitive resources; are sensitive to the social surrounding; and are influenced by their cultural mind-set (Rodríguez-Priego and van Bavel 2014). Behavioural economists have shed light on the need to account for these behavioural elements of the decision context, beyond simple profit maximisation, in order to predict economic agents' responses to different policy instruments. Also in the farming sector, empirical research has shown that farmers' decisions are motivated by many factors that cannot be directly observed. Accounting for these behavioural factors in policy evaluation is particularly important in two cases (van Bavel et al. 2013):

First, when behavioural change is the main objective of the policy. For example, getting farmers to select appropriate crop insurance, contract agri-environmental measures or adopt organic farming practices.

Second, when agent's behavioural response affects the effectiveness of a given policy (even if the aim of the policy itself may not be to change behaviour). For example, policies intended to support farm income do not seek to change farmers' behaviours, but the support may impact farmers' decisions, even with decoupling of direct payments, and this is likely to impact farm income in an unforeseen way.

Some of these behavioural factors (such as risk aversion, time inconsistent behaviour, intrinsic motivations and social norms) and examples in the farming context are described in Table 2.

While intention surveys ("what would you do if") could give an indication of farmers' reactions in the context of behavioural factors, experiments collecting stated preferences in a structured way (as is the case for DCEs) or lab and field experiments can reveal which of these factors are relevant to explain specific farmers' decisions and what types of policies can work in the presence of these behavioural factors (Prendergrast et al. 2008).

It is possible to improve the effectiveness of the agricultural policy by taking into account such behavioural biases in their design and implementation. For example, if social norms are a key driver of the adoption of new agricultural practices, supporting community engagement is likely to be more efficient than setting up individual payment schemes to encourage adoption.

> « Predicted welfare effects from policy changes are always uncertain, in part because of imprecisely known parameter values in the policy simulation models used. » (Harrison, Lau, and Rutström 2011)

Finally, experiments can reveal the value of behavioural parameters, which can then be used in simulation models (instead of calculating the value of such parameters through calibration).

[7] As will be discussed in more detail in Chapter 4, the possibility of identifying the true causal effect for the sample of participants in the experiment is called the 'internal validity' of the experiment.

2.2.3 Test policies prior to implementation

Experimental approaches encourage moving beyond a status quo in evidence-based policy making.

> « Evidence can sometimes only reflect what worked in the past, or in other countries. Focusing on what seemed to work in the past reinforces the status quo (…), while we need to strive for new alternatives and new knowledge to meet current challenges » (Breckon 2015).

Reliance on past observations can indicate that a certain policy did not work, but it provides little information on what new policies would do better. Especially when farmers' reactions are expected to be impacted structurally, yet in an unknown way, by the new policy, information on the past is often insufficient. Moreover, in many cases the naturally occurring variation in policy treatments is too small to allow economists to infer reliably the effects of new policies. For example, the differences in CAP direct payments that can be observed across farmers are likely too small to understand what would happen in case the current CAP payment scheme would undergo a substantial restructuring. Especially for innovative policies, that may be very different from currently implemented policies, we may know little on how stakeholders will react. Lessons based on past experiences may then be of little value. In such cases, experimental approaches offer the possibility of testing new policies ex-ante on a small scale.

> « Lessons about policy design are vastly cheaper when learned in the laboratory than through trial and error in the real world. » (Reeson & Nolles, 2009)

Experiments allow testing a new policy in a virtual environment, or at small scale, before full-scale implementation, so as to limit the political and economic cost of a trial that could fail.

> « The advantage of an experimental approach is that you can be innovative and cautious at the same time…try things out in an overly tentative manner. » (Berk et al. 1985).

Experimental approaches also offer the possibility of rapidly testing slight variations of a policy. For example, using an experiment one could test whether formulating a payment as a "compensation for the additional costs associated with an agri-environmental measure" or a "reward for contributing to the environmental conservation" makes a difference.

Designing experiments for ex-ante evaluation can also contribute to provide new ideas to policy-makers, for example by shedding light on the details needed for effective policy implementation. Experiments are helpful in identifying potential design flaws, especially when there is no empirical evidence or theoretical predictions are difficult if not impossible to derive (Normann and Ricciuti 2009).

The experience described by Cummings et al. (2004) offers a glimpse of how the design of experiments for ex-ante evaluation can impact policy-making. They explain how the joint work on the design of a conservation auction allowed legislators to realise the complexities of implementing an auction-like process, and the vast array of auction mechanisms that could be used, which they had not imagined with the institutional constraints and political considerations they had in mind. Moreover, the behaviour of the participants during the experiments and feedback received after the experiments were used to fine-tune the

auction procedures. Also the development of easily understood instructions for participants of the experiment provided a useful framework for the instructions that were sent to farmers when the auction was effectively implemented.

Table 2: Some examples of behavioural elements important to take into account in agricultural policy design and evaluation

Behavioural elements	Relevance in the agricultural context
INTERNAL FACTORS: HABIT AND COGNITIVE PROCESSES	
Risk aversion Reluctance of a person to accept a bargain with an uncertain payoff rather than another bargain with more certain but possibly lower expected payoff.	With a sample of 107 French farmers, Bocquého et al. (2013) have shown that more dedicated, wealthier farmers are less risk averse and that a large proportion is averse to losses and overestimate low probability extreme-outcome events.
Loss aversion Individuals are averse to loss, i.e. they weight the value of losses more than the value of gains (Kahneman 2003).	Loss aversion (Kahneman 2003) can explain why some farmers are reluctant to contract an agri-environmental measure on input use reduction, even if the reduced costs of input use could compensate yield loss: farmers are more sensitive to the yield losses induced by the adoption of an agro environmental contract, than to the increase in their expected revenue. As a result, an agri-environmental contract highlighting that the payment is a compensation to cover the implementation costs and revenue foregone from a change in practices is potentially less efficient than a contract stating that the payment is rewarding the provision of environmental services (Grolleau et al. 2015). Given loss aversion, it has also been shown that penalties are more efficient than bonuses to encourage the adoption of new milking practices among dairy farmers to improve cattle health (Valeeva et al. 2007; Huijps et al. 2010).
Ambiguity aversion Preference for known risks over unknown risks. Ambiguity averse agents avoid options where incomplete information makes the choice feel more risky.	Farmers are usually expected to choose the technology that provides the highest expected utility conditional on their aversion to risk. But, new technologies often involve ambiguity such that the probabilities of different outcomes are not known. Lack of information (or lack of confidence in this information) could then explain why some technologies are received less favourably than others (Ross et al. 2012). The quality and credibility of information policies are therefore crucial in uncertain contexts. For example, Bradford et al. (2014) found that the adoption of genetically modified soy seeds is not so much related to risk aversion, but rather to ambiguity-aversion. GM corn containing an insect-resistance trait which reduces the ambiguity of pest damages, therefore leads to a important speed-up in adoption.

Time inconsistency

People's preferences tend to change over time, such that decisions which are made today can be inconsistent with long term preferences.

Bocquého et al. (2013) observed time preferences of French farmers and found that the elicited individual discount rate varies with the time delay, which provides some evidence of possible reversal effects in long term time preferences. They also found that time preferences depend on the time-lag between payments. These results call for the use of more complex models of inter-temporal choices to design voluntary measures. In particular, it should be taken into account that the costs in the short run may be overvalued compared to the long run benefits by farmers (Mzoughi et al. 2014). This leads to the difficulty of engaging today in investments or actions that involve pay-offs in the future, even if these choices would be optimal from a long term perspective.

Status Quo bias/Default bias

Tendency to prefer things to stay the same, selecting the default option where available.

The framing, design and presentation of choices, and the default options behind these choices, can be altered to try and help individuals to make the optimal choice (Thaler and Sunstein 2008; Choi et al. 2003). For example, farmer cooperatives may be influential in affecting "default options" such as providing pest-resistant seeds as a default input to its members. Such approaches still allow farmers to opt-out should they prefer a traditional technology, but evidence shows that such defaults have a significant impact on behaviour.

Choice overload

The phenomenon occurs when many equivalent choices are available. Making a decision becomes overwhelming due to the many potential outcomes and risks that may result from making the wrong choice.

Information overload tends to modify decision making processes. For example, people tend to focus on only one or two choices even when there are several possible outcomes (attentional bias). They may also focus on the impact of individual choices rather than the consequences of many choices taken together (choice bracketing). With agricultural production systems continuing to become more complex and information intensive, demands on farm decision-makers in acquiring, evaluating, and processing of information are ever increasing. It makes increasingly important that farmers be given quality information, delivered in the proper format, and at the right time.

The importance of intrinsic motivation Individuals' choices are influenced by extrinsic incentives (such as financial incentives), but also intrinsic ones.	Policy focusing on economic incentives could have opposite effects if individuals feel their intrinsic motivations are ignored. Economic incentives can crowd out intrinsic motivations and overall reduce agents' effort. For example, Greiner and Gregg (2011) found that graziers had a very high level of conservation and lifestyle motivation and were motivated to lesser extents by financial/economic considerations. Conservation programmes need to take advantage of farmers' stewardship ethic for maximum effectiveness and efficiency, and minimise the risk of crowding out intrinsic motivation and altruistic behaviours.
SOCIAL FACTORS: OTHER-REGARDING BEHAVIOUR AND SOCIAL NORMS	
Reciprocity and fairness People may gain satisfaction from the levels of satisfaction and behaviours of others. (Fehr and Schmidt 1999). One explanation can be the positive emotional feeling (warm glow) people get from helping others (Andreoni 1989).	Individuals are most likely to adopt farming practices that generate positive externalities on others (Kelsey 2013). Investments in communication to the general public on the on-going greening of farming practices is likely to further encourage farmers.
Social norms and positional concerns Tendency to base actions and beliefs on what others are doing or believing. The resulting "bandwagon effect" refers to the probability of individual adoption increasing with respect to the proportion who have already done so.	Moser and Barrett (2006) observed that the desire to conform to the behaviours of other farmers appears to have a positive effect on agricultural technology adoption, beyond the effect of social learning. Chen et al. (2009) have shown that Chinese farmers are more likely to renew an agri-environmental contract if they know their neighbours do so as well. Kuhfuss et al. (2014) found that vine-growers are more willing to provide environmental efforts when they have given some indications that their neighbours will also do so.

3 Economic experiments for agricultural policy evaluation: a review of the literature

This chapter reviews the most noteworthy applications of economic experiments in the field of agricultural policy design and evaluation – without laying claim to be complete. In order to structure this overview of the literature, the following classification criteria were used:

- **Type of economic experiment:** discrete choice experiments; laboratory experiments; field experiments; randomised controlled trials.
- **Type of policy to be evaluated:** policies aiming at supporting farm incomes; policies to enhance the provision of environmental services by farmers; policies supporting farm modernisation and agricultural innovation.
- **Purpose of the experiment:** to measure the net impact of existing policies; to test new policies prior to implementation; to assess the role of behavioural factors in farmers' responses to policies.
- **Participants in the experiment:** students, farmers, consumers or citizens.
- **Country in which the experiment was conducted:** European Union; non-EU industrialised country; developing country.

In what follows, we classify the literature first by the type of experiment. We then group experiments broadly around the types of policies they evaluate.

3.1 Discrete Choice Experiments

The most widely used experimental approach to evaluate agricultural policies is the discrete choice format. The majority of these studies have focussed on estimating the benefits of local or regional agri-environmental practices in the EU and non-EU industrialised countries. The classic example of this type of study is a DCE with members of the public (e.g. landscape users or citizens in general) to assess the economic benefits of policy-induced changes in land use or environmental practices.

Only in recent years has the DCE method been applied on the producer side. Examples include assessments of farmers' prospective responses to new policies, estimation of the farm-level costs of participation in agri-environmental schemes, or analysis of the role of non-monetary factors as drivers of farmers' responses to policies.

By bringing supply-side and demand-side valuations together, DCEs allow comparing the benefits and costs so as to maximise net benefits. Surprisingly, there are hitherto no DCE studies which have aimed to assess demand and supply for the same environmental good simultaneously. In this section, we focus mainly on DCE applications on the producer side. We only

briefly discuss the literature on DCEs on the consumer side, focusing on two examples of experiments to assess public preferences for environmental improvements.

3.1.1 Assessing farmers' preferences for agri-environmental schemes

Most DCEs with farmers have focused on eliciting farmers' preferences for participation in agri-environmental schemes. They reveal farmers' willingness to accept (WTA) compensation for participation in such schemes. They thus provide an estimate of the cost of compliance with such agri-environmental contracts. Most interestingly, these experiments allow assessing the costs of specific attributes of the schemes and the heterogeneity in farmers' preferences and responses. As such, they provide policy makers with information on which characteristics of the contract are considered most problematic or most desirable from farmers' point of view, and they provide information on which types of farmers are most likely to participate, allowing to better fine tune and target policies.

Espinosa-Goded et al. (2010) investigate the factors affecting farmer's willingness to participate in an agri-environmental scheme paying Spanish farmers to cultivate alfalfa (a nitrogen-fixing crop). Data were obtained from a survey of 300 farmers undertaken in two regions of Spain - Aragon and Andalusia, representing different farming systems. The experimental design was based on a scheme which had been proposed for implementation in both Aragon's and Andalusia's Rural Development Program (RDP) for 2007–2013. Farmers were confronted with choice sets in which they had to choose among alternative contracts, each involving distinct implementation requirements and a given compensation payment. The results indicate that free choice of the land offered for the programme and unrestricted use of the alfalfa crop significantly increase respondents' willingness to sign a contract.

Christensen et al. (2011) conduct a DCE with 444 Danish farmers to assess their willingness to sign contracts for pesticide-free buffer zones. Their research was motivated by the observation that Danish farmers had been less interested in agri-environmental schemes than anticipated. As in the study of Espinosa-Goded et al. (2010), respondents were asked to choose among alternative contracts, each with different requirements and payment levels. The results show that farmers prefer contracts with a flexible buffer zone width, shorter contract periods, greater flexibility in fertilizer use, and the option to quit the contract from year to year. Especially this last option is very highly valued, indicating that farmers feel very strongly about this possibility to re-evaluate their participation into the contract at a regular basis. The authors conclude that the overall flexibility of the contract might be more important to farmers than the land use restrictions imposed. The experiment yielded another important piece of information for policy makers: 14% of respondents consistently chose 'no contract' (i.e. business as usual) which clearly indicates that these farmers might be difficult to motivate to enrol in agri-environmental schemes.

Ruto and Garrod (2009) gathered data from DCEs with farmers in 10 case study areas across the EU. The authors confront their respondents with alternative designs of a hypothetical agri-environmental scheme, each characterised by five attributes: compensation payment, contract length, whole farm or partial area participation, degree of flexibility in implementing conservation practices, and administrative burden for the farmer. In general, farmers were found to require greater financial

incentives to join schemes with longer contracts or that offer less flexibility or demand more paperwork. In addition, a large segment of farmers ('low resistance adopters') is found to be willing to accept relatively small incentive payments for their participation in schemes, even when these schemes are offering relatively little flexibility and high levels of additional paperwork, in contrast to a segment of 'high resistance adopters' who would be difficult to motivate even with very flexible schemes and high financial compensation. This suggests that the targeting of schemes to the most interested farmers may provide large efficiency gains.

Breustedt et al. (2013 a; 2013 b) use data from a DCE on agri-environmental schemes aiming to protect ground-nesting birds in permanent grassland in two grassland regions of Schleswig-Holstein, Germany. As in the studies reviewed above, farmers were asked to make a discrete choice among three alternative contracts (each with different land use restrictions and payment rates) or a 'no participation' option. The contractual attributes related to stocking rates, fertilizer usage, set mowing dates, contract length and payment rates. Breustedt et al. (2013 a) find that, as expected, higher payment rates increase farmers' willingness to sign an agri-environmental contract whereas stricter management prescriptions reduce participation. The estimates of marginal willingness-to-accept show that farmers consider contractual obligations relating to fertilizer use and mowing dates to be particularly demanding. Besides contract design variables, a number of farm-specific variables affect respondents' willingness to participate. Farmers, for instance, with previous experience in agri-environmental schemes are more likely to engage in a contract than those without such experience. Farms with high land use intensities request higher compensation payments than less intensive farms.

Breustedt et al. (2013 b) investigate not only which among alternative agri-environmental contracts is chosen, but also how much land farmers would put under contract. For contract 1, farmers had to enrol at least 5% of their permanent pasture; contract 2 required the enrolment of at least 10% and contract 3 of at least 20% of a farm's permanent pasture. This extension to the experimental design enabled the authors to provide information of direct relevance to policy makers, such as the estimation of a supply curve for land offered under conservation management as a function of the payment rate. More importantly, it allowed the authors to formulate optimal contracts (in terms of management prescriptions and payment levels), such that environmental benefit is maximised for a set of alternative program budgets.

Birol et al.'s (2006) DCE investigates Hungarian farmers' preferences for undertaking traditional farming methods on their small farms or home gardens, including crop variety diversity; cultivation of landraces; traditional integration of crop and livestock production; and use of organic production practices. The results reveal that farmers located in the most isolated communities derive the highest value from crop diversity, and among those, elderly derive the highest value from traditional landraces. Given the environmental, cultural and historical benefits that home gardens generate, the authors suggest that they should be supported by means of an appropriate agri-environmental scheme. In addition, the authors were able to identify those smallholder households that value the benefits the most and suggest that the support schemes be targeted on those households.

Kuhfuss et al. (2014) conduct a DCE with 317 wine growers in the South East of France to test whether the introduction of a collective dimension in agri-environmental contracts can entice more farmers into agri-environmental schemes, thereby initiating group dynamics towards reduced use of pesticides on the land. Their collective dimension takes the form of a monetary bonus paid per hectare to farmers participating in the contracts, in addition to standard individual compensation, on the condition that a predefined collective enrolment target (50% of the area of interest) is reached. Respondents were asked to choose between pairs of weeding regimes, each characterised by a set of five attributes, one of which was whether a collective bonus was offered or not. The others related to herbicides use restrictions and whether administrative and technical assistance was offered. The results provide clear evidence that introducing a collective bonus can improve participation rates, increasing the proportion of land enrolled by each farmer and lowering the individual WTA.

Greiner (2015) reports the results of a DCE conducted with 104 pastoralists and graziers in north Australia. The experiment was designed to explore the factors affecting respondents' willingness to sign voluntary on-farm biodiversity conservation contracts. Each choice set consisted of three contract alternatives and a 'none' option. Contract attributes, i.e. conservation requirements, stewardship payment, contract length and flexibility in contract conditions, all significantly influence choices, but with significant heterogeneity across respondents. Respondents require a greater monetary incentive to sign up to longer contract periods or alternatives causing higher opportunity costs. Introducing some level of contract flexibility significantly and positively influences contract adoption. Financial aspects of the contracts were the most important for about half of the pastoralists interviewed, while willingness to participate of the other half was significantly influenced by their attitudes and non-monetary motives – factors which can only be identified with a DCE.

Broch and Vedel (2012) use a DCE to study the heterogeneity of farmers' preferences for afforestation contracts as a means of targeting contracts on the specific needs of individual landholders. Four attributes are investigated: purpose of afforestation, option of cancelling the contract, compliance monitoring, and payment level. All attributes present a potential conflict between farmers' and authorities' interests. Results reveal that the option to cancel the contract decreases farmers' required compensation level, whereas external monitoring increases it. Moreover, farmers are willing to accept a lower compensation when the purpose of afforestation is to protect biodiversity and ground water rather than recreation. This indicates that farmers have preferences not only for the characteristics of agri-environmental contracts but also for the objectives of the policy scheme. There is heterogeneity in farmers' preferences: for example, farmers who already have forest land do not find the option of cancelling the contract important, whereas farmers who rely on the farmland for income require higher compensation if this option is not given. The findings indicate potential for efficiency gains from targeting the contracts at specific groups of farmers.

The study by Beharry-Borg et al. (2013) applies DCE methodology to assess the use of incentive payments for farmers to adjust their land management practices in order to protect water quality. In contrast to the previous studies, the payments would be paid by a water company instead of being part of European or national policies. The

authors quantify the financial compensation (WTA) required by farmers to adopt water quality enhancing agricultural land management practices in two watersheds in Yorkshire, UK. Heterogeneity in farmers' preferences and compensation requirements is found to be related to the emphasis on sheep or cattle/dairy production within mixed farming businesses in the study area.

These studies reveal estimates of farmers' willingness to accept to participate in such schemes, i.e. the minimum compensation required in exchange of participation. They thus provide an estimate of the cost of compliance with such agri-environmental contracts. More importantly, the DCEs allow to decompose the total cost of compliance into specific part-costs for each attribute of the schemes, and to reveal the heterogeneity in farmers' preferences and WTA. As such, they provide policy makers with information as to which characteristics of the contract are considered most problematic or most desirable from the farmers' point of view, and they provide information on which types of farmers are most likely to participate, allowing to better fine tune and target policies. Although the precise results are obviously specific to the scheme and location studied, the general conclusion from these studies is that introducing some flexibility in the contract considerably increases farmers' willingness to participate, as does shorter or more flexible contract duration.

Overall, DCEs seem to be particularly useful for optimising the design of voluntary agri-environmental schemes. If the policy designer knows beforehand which contractual obligations farmers consider to be particularly demanding, this information can be taken into account in designing the contracts and calibrating payment rates. In this way, the informational imbalance between the policy designer and the farmer, which often impedes the design of efficient conservation contracts, can be mitigated. An outstanding feature of DCEs in this respect is that they provide estimates of farmers' participation costs differentiated by farm type, location or resource settings, thereby allowing the policy maker to devise contracts for different types of farmers and different resource settings. This is likely to boost the cost-effectiveness of such schemes compared to the standard one-size-fits-all contract.

3.1.2 Assessing farmers' responses to first-pillar policies

Schulz et al. (2014) is one of the few applications of a DCE to inform the design of CAP first-pillar policies. The authors explore farmers' prospective responses to the new 'greening' provisions of the Common Agricultural Policy. They aim to investigate how the 'greening' provisions are likely to affect farmers' willingness to comply; to assess the farm-level costs of complying with the 'greening' provisions; and to explore the heterogeneity of preferences and costs among farmers. Their analysis is based on a DCE with 128 German farmers conducted in the summer of 2012, i.e. before the policy entered into force. Participants were asked to choose between a 'greening' option with a given set of management prescriptions and an 'opt-out' alternative with a stipulated cut of the single direct payment.

The authors find that farmers perceive 'greening' as a costly constraint, but not all farmers are equally affected and not all 'greening' provisions are regarded as equally demanding. The share of Ecological Focus Area (EFA) has a strong impact on the choice to comply with greening provisions. The surface of EFA is valued at the rental value of arable land. If productive uses of the EFA (by opposition to EFA must be set-

aside) are allowed, the likelihood of greening being preferred increases. By contrast, and maybe surprisingly, the possibility of counting as EFA existing landscape features on a farm (such as hedges or trees) or land enrolled in agri-environmental schemes does not affect the likelihood of 'greening' being preferred to 'opt-out'. Farms with high opportunity costs of arable land are found to be more inclined to opt out and voluntarily forgo the 'greening' premium than farmers with lower land opportunity costs. The analysis also revealed that a small group of farmers will fiercely reject 'greening': 14% of respondents never chose a 'greening' alternative.

The research by Lips and Gazzarin (2008) represents another application of a DCE to inform first-pillar type policies. The authors elicit the intentions and preferences of Swiss dairy farmers for pursuing their work as farmers after abolishment of the milk quota in Switzerland. The analysis is based on a DCE with 304 dairy farmers carried out in the Eastern part of Switzerland before the quota was abolished. Respondents were asked to choose between the status quo (milk production) and alternatives consisting of several combinations of four attributes. The attributes comprise the type of farming (milk production, suckler cows, farming without livestock, leaving farming), employment status (self-employed, employed), days of holiday per year, and income per year. The results indicate that there is a strong preference to stay in dairy production. In order to entice farmers to give up milk production, a financial compensation (in the form of additional income) of at least CHF 25'000 per year would be necessary. This is more than 60 percent of the income per family member in dairy production. Given this very high compensation requirement, which is partly driven by non-monetary motives, the authors conclude that an agricultural policy programme based on financial incentives designed to motivate dairy farmers to quit milk production is likely to fail.

The study by Paulrud and Laitila (2010) is an example of a DCE designed to inform Swedish energy crops policy. The authors use the DCE method to analyse how Swedish farmers assess the relative merits of the characteristics associated with growing energy crops. An additional goal is to explore farmers' willingness to grow energy crops as a function of different levels of income and subsidies. The authors conducted two DCEs. In the first experiment, farmers had to choose between two (unspecified) energy crops, each characterized by six attributes. The attributes related to the institutional and agronomic setup of the production process, e.g. how the crops would be integrated into arable rotations, whether they would be grown independently or on contract, whether conventional farming machinery could be used in their cultivation, whether a subsidy is offered and whether the crop would affect the visual quality of the countryside. The second experiment specified four energy crops (willow, hemp, energy grains and canary grass) and characterised them in terms of their net contribution to farm income with and without subsidy. Respondents were then asked to state how many hectares of each crop they would be willing to grow on their arable land with and without energy crop subsidy. The results of the first experiment suggest that the visual impact on the countryside and the rotation period of the energy crops have a significant impact on the willingness to cultivate such crops. The second experiment showed that an increased contribution of a specific crop to net income raises the amount of arable land devoted to growing that crop. It was also found that farmers' age, farm size and the geographical location of the farm had a significant effect on the willingness to grow

energy crops, whereas other farm features (such as share of leased land or set-aside land or farm type) did not.

3.1.3 Assessing public preferences for environmental improvements

DCEs focussing on the estimation of the benefits for the public of the provision of environmental services by agriculture have covered a wide range of issues. We provide here only two examples of how DCEs have been used to inform the design of agri-environmental policies by providing policy makers with estimates of the public's preferences for environmental improvements. Many more examples can be found in the literature on the evaluation of public preferences for environmental, cultural and social features of landscapes, natural habitats or species. The book "Choice Experiments Informing Environmental Policy" edited by Birol and Koundouri (2008) provides a good account of the literature up until 2008.

Hanley et al.'s (1998) work focuses on benefit estimation for Environmentally Sensitive Areas (ESA) in Scotland, the first generation of co-funded agri-environmental schemes in the UK. For a specific ESA in Scotland, the authors identify farm woodlands, archaeological features, heather moors, wet grasslands and drystone walls as important landscape features, which could be affected by the ESA. Their analysis shows that the public places the highest value on schemes that improve farm woodlands, followed by heather moors and wet grasslands. The lowest public values were found for schemes that target the improvement of archaeological features. The authors conclude that the ESA should therefore prioritise the provision of woodlands on farms.

Another DCE study that focuses on nature conservation from public perspective is that by Li et al. (2004). The study investigates the preferences of the Finnish public for increasing the area under preservation in Natura 2000 areas. The analysis of the data reveals considerable preference heterogeneity among the public. Overall, the Finnish public is more sensitive to a decrease in the Natura 2000 area than to an increase of that area: the mean willingness-to-pay (WTP) to prevent a decrease was found to be four times higher than the mean WTP for an increase. Moreover, the marginal value of nature preservation falls to zero after an increase of only 3% in the size of the currently preserved area. These results are expected to aid policy makers in their cost-benefit analysis of alternative nature preservation programmes in Finland.

3.2 Lab and field experiments

Several studies have used lab and field experiments to analyse the best design for the agricultural policy. A small number of studies have focused on first-pillar type of agricultural policies, both in the US and in the EU, while the literature on experiments for agri-environmental policy evaluation is more extensive.

3.2.1 Applications to assess the net impact of first-pillar type policies

Most studies related to first-pillar type of agricultural policies focus on the design of support payments and their impact on production decisions.

Bahrs et al. (2008) investigate different trading mechanisms for EU decoupled payment entitlements introduced with the Fischler Reform of the CAP in 2005. Since, at the time, policies had changed only

recently and the trading had just started, real-world market evidence was spurious and anecdotal. Thus, controlled laboratory experiments were conducted to provide first insights into potential market outcomes and how they are influenced by trading rules. The authors assume that, at least in early stages, the exchange of subsidy entitlements would occur in a private and decentralized setting - for example, between neighbours in the same village. Their experimental setup reflected this situation by allowing student subjects to trade with only one other subject at a time. They compare this setting of decentralised bilateral negotiations with a setting with centralised trading, such as posted-bid or posted-ask trading.

At the start of the experiment, each subject was endowed with land (defined to be untradeable) and tradable 'high-value' and 'low-value' subsidy entitlements. Ten rounds were played with 10 players. At the end of each round, subsidies were paid only for 'activated' entitlements, i.e. only if the participants had the corresponding land. The main finding is that market institutions have a large impact on outcomes: decentralised bilateral negotiations do not generate outcomes that are as efficient as those in the more centralised market institutions.

Motivated by policy changes introduced in the US 2002 Farm Act, McIntosh et al. (2007) investigate supply responses to countercyclical payments given to farmers (in addition to direct payments) in a world of price uncertainty. Students first had to allocate a fixed number of acres to a base crop, i.e. a crop eligible to enrol in the Direct and Countercyclical Program, and a non-base crop. Each participants' task was then to allocate acres under three different policy scenarios: (a) the baseline case: price uncertainty with DPs and without countercyclical payments (CCPs); (b) the CCP case: price uncertainty with DPs and the CCPs; (c) the policy risk case: participants had to make their allocation decisions while the type of policy was uncertain (it could be the baseline, the CCP case or another scheme called 'mandatory base updating').[8] Ten rounds were played for each of the 3 scenarios, with 88 student participants. The experiment allowed comparing behaviour in the CCP case and the policy risk case against the baseline case to better understand how CCPs and policy risk affect crop allocation decisions. The experimental evidence suggests that CCPs lead to greater income certainty but less efficient production decisions and (possibly) higher government payments because participants allocated more acres to the base crops relative to the baseline, which are eligible for countercyclical payments - supporting some of the criticisms of CCPs.

Nagler et al. (2009) and Bastian et al. (2008) report the results of an economic experiment to investigate the impacts on production decisions of the bond scheme, an alternative to the single farm payment (SFP) scheme. Producers would receive a buy-out of future SFP entitlement, and be issued with bonds. Such a bond could consist of annual payments for several years, or to a one-time lump-sum payment equivalent to these annual payments. The bonds would be transferable and could be sold. Hence, this would correspond to a guaranteed stream of income during a transition period of 15 to 25 years in exchange for giving up all future subsidy entitlements. In a lab experiment with economic students, the

[8] The 2002 Farm Bill allowed farmers who received direct payments between 1996 and 2002 to choose between keeping their old base acreage or updating base acres to reflect average planted acres for eligible commodities during the 1998-2001 crop years. In the experiment, mandatory base acreage updating meant automatically adjusting the base acreage to the cropping allocations of the current round.

authors investigate four experimental treatments each involving 40 trading periods (representing production seasons): (1) the base treatment of no policy (NP) intervention for 40 trading periods; (2) a coupled price-support policy for 40 trading periods; (3) a coupled price-support policy for the first 20 periods, followed by a decoupled annual subsidy for the last 20 trading periods; and (4) a coupled price-support policy for the first 20 periods followed by a buyout bond consisting of a decoupled lump sum subsidy received by participants in period 21, with no policy intervention for the remaining 19 periods.

As expected, production with a coupled price-support policy was substantially higher than with no policy. A shift from price support to equivalent decoupled annual or lump-sum payments resulted in production levels at or near no-policy levels, providing evidence to support the theoretical prediction that bond schemes would not result in production distortions.

Nagler et al. (2013) conducted a lab experiment with both students and agricultural professionals to estimate the capitalisation of subsidies into land rental prices. For both types of participants to the experiment, land rental prices converged to a value approximately four and a half tokens higher when a 20-token per-unit subsidy was paid, compare to a treatment with no subsidy. These experimental results are similar to empirical estimates: one fourth to one fifth of the subsidy is passed on to the rental price of land.

A last example concerns the effect of price support. Maart-Noelck et al. (2013) study the impact of a price floor on investment behaviour. Price floors such as intervention prices for agricultural commodities or guaranteed feed-in tariffs for renewable energies are often claimed to have a stimulating effect on investment behaviour because they reduce the risk of unfavourable outcomes. This study uses an experimental approach to test whether this is indeed the case. The study design is based on two scenarios: one 'with price floor' (WPF) and one with 'no price floor' (NPF). Participants were 101 German students of agriculture. A first group of students faced at first the NPF scenario without any market intervention. After ten repetitions the participants played the WPF scenario. The other group faced both treatments in reverse order.

The main results derived from the experiment can be summarised as follows: First, actual investment behaviour did not differ significantly with respect to the presence of a price floor. Second, abolishing the price floor, leads to investment inertia: participants who first played the WPF treatment tended to invest more reluctantly over both treatments than participants who faced the treatments in reverse order. Third, participants invested, on average, much later than theoretically predicted by the net present value (NPV) approach and much earlier than predicted by the real options approach (ROA) in either treatment. Fourth, participants learned from personal experience during the experiment and approached the theoretical ROA benchmarks over time. The authors conclude that price floors do not necessarily stimulate investment behaviour - although higher levels of price floors than those used in this experiment may still be found to be effective - and that investment theory does not predict actual investment behaviour particularly well.

3.2.2 Applications to agri-environmental policies

Recently, a number of studies have analysed experimentally the potential use of conservation auctions (or conservation

tenders as they are usually called in the literature) in agri-environmental policy. Economic experiments involving auctions are widely used in the traditional areas of competition policy and auction design, but agricultural and agri-environmental policy experiments are relatively new.

Environmental auctions are a mechanism whereby conservation contracts with well specified sets of management prescriptions are allocated to private landholders on the basis of competitive bidding. Bids take the form of requests for financial compensation in return for implementing the contractual obligations on one's land. Auctions are considered as a more efficient alternative to the currently used fixed-price contracts of the CAP's agri-environmental schemes, given the potential to acquire a greater quantity of conservation effort for the same budget (Latacz-Lohmann and van der Hamsvoort 1997).

While such auctions have been organised in the US or in Australia, to date, no EU country has made use of conservation tenders at a large scale, although the legal requirements for their use have been created.

The experimental study of conservation tenders analyses what type of auctions, what information provision etc. generate the best outcomes. They use either controlled lab experiments or the sequential combination of lab experiments and small-scale field trials.

Cason et al. (2003) investigated the BushTender trials in the Australian state of Victoria, which were designed to conserve the last remaining patches of native vegetation by fencing these areas off, excluding livestock, refraining from using pesticides, etc. Farmers who have native vegetation on their land were invited to submit financial bids for agreeing to conserve the land in pristine state. This study was the first to propose a combination of lab and field experiments: prior to testing this programme in the field, certain design problems, in particular the amount and choice of the information to be communicated to landholders before the bidding session, was investigated experimentally in the lab with students. In a second stage, the experiment was moved to the field, with farmers as participants.

They tested whether or not it is a good idea to reveal the environmental benefit that the contract is expected to deliver before the landholders are submitting their bids. Revealing the benefits may provide landholders with information on the most beneficial land use changes or what types of land to offer for conservation. On the other hand, revealing information may lead to strategic bidding behaviour. The authors find that revealing the benefits led to lower environmental benefit and higher landholder profits.

In follow-up work, Cason and Gangadharan (2005) study the outcome properties of uniform versus discriminatory-price auctions for reducing non-point source pollution resulting from fertilizer use. In the uniform-price format, all successful bidders are paid the price of the lowest rejected bid, while a discriminatory-price auction pays successful bidders their own bid. The results from their lab experiment with students show that overbidding is more pronounced in the discriminatory-price auction but it had superior overall performance in terms of value for money for the environmental agency. This can be explained by the fact that, in the uniform-price auction, some landowners are 'overpaid' because, by construction of the auction, they receive payments largely in excess of their opportunity cost.

Schilizzi and Latacz-Lohmann (2013) conducted a framed lab experiment with students to assess the performance of budget-constrained versus target-constrained conservation tenders and compare this to equivalent fixed-price agri-environmental schemes. In the budget-constrained format, the agency accepts bids until a given budget is exhausted, while in a target-constrained auction the agency sets an overall environmental target and accepts bids until this target is achieved. The study was framed in the context of an agri-environmental scheme for the adoption of farm technologies to reduce non-point source pollution. The experiment involved multiple bidding rounds for both auction formats. The results indicate that the choice of format did not have much effect on auction performance.

The Flint River Drought Protection Act passed in Georgia in 2000, mandated that an auction-like process for foregoing irrigation be implemented in years of drought. By running experimental auctions, both in the laboratory and with farmers, prior to the actual irrigation auctions, Cummings et al. (2004) were able to make suggestions on the auction design to make it more competitive and at the same time better understood by the participants. They compared uniform and discriminative pricing formats, as well as different tie-breaking rules[9] (inclusive or random) to inhibit collusion, and whether or not information was provided on the highest accepted price of offers. The aim was to maximise the land area taken out of irrigation, given a fixed budget. The random tie-breaking rule, and not announcing the price of the provisionally accepted offers, resulted on average in better value for money for the procuring agency. These two features, along with the discriminative-price format, which performed similarly to the uniform-price format in the experiments but was preferred for political reasons, were adopted in the actual field auction conducted in 2001.

Another question investigated with experiments is the monitoring of farmers' compliance with the prescriptions of conservation contracts. Kawasaki et al. (2012) have analysed the impact of imperfect monitoring, which provides farmers with a moral hazard incentive for non-compliance because they can receive conservation payments without implementing the conservation scheme. The authors compare bidding behaviour and auction performance for discriminatory-price and uniform-price auctions in an imperfect monitoring environment. While their theoretical analysis suggests that there should be no performance differences between the two auction formats, their laboratory experiments revealed contradictory results: while the discriminatory-price auction had an advantage in terms of cost-effectiveness, the uniform-price auction tended to have higher overall efficiency when non-compliance behaviour was taken into account. The analysis thus highlights the need for policy makers to consider the efficiency of policies holistically, not just at the moment of allocating contracts.

Another key aspect of conservation contract design is that environmentally friendly land use practices often deliver greater biodiversity and habitat enhancements when they are located on spatially adjacent land parcels with connections between them. One approach to spatially coordinating conservation efforts across multiple private properties is the agglomeration bonus, first proposed by

[9] A tie-breaking rule must be applied when two or more bids are equal at the cut-off price. The tie-breaking rule determines which of the bids is accepted and which are rejected.

Parkhurst et al. (2002) and Parkhurst and Shogren (2007). By rewarding similar land uses on adjacent parcels, the agglomeration bonus provides economic incentives for the creation of non-fragmented land use patterns on the landscape. Because of the complexity of this incentive mechanism, theoretical modelling can only yield limited insights into its environmental effectiveness. Parkhurst et al. (2002) thus tested the agglomeration bonus experimentally. Their experiments (with students) indicate that including an agglomeration bonus significantly decreased the fragmentation of the conserved land compared. Allowing for communication between participants before the start of the game further improved results. In later work, Parkhurst and Shogren (2007, 2008) check the robustness of their results in a more realistic, but more complicated, spatial-coordination setting. The agglomeration bonus remained an effective policy tool in this more complex setting once participants gain experience. The studies by Reeson and Tisdell (2010), Rolfe et al. (2009) and Banerjee et al. (2014) link the contract design and conservation auction literatures by integrating an agglomeration bonus into a conservation auction and analysing the cost-efficiency of different auction formats.

Latacz-Lohmann et al. (2012) evaluate the efficiency of a new contract design, in which payments are linked to environmental outcomes (rather than management prescriptions) and contracts are tendered (rather than paying landholders uniform prices). Controlled lab experiments with students were run, systematically varying the rate at which payments are linked to environmental outcomes. Results show that when the share of the payment linked to outcome increases, more effort in conservation is observed but the participation rate falls. Indeed, outcome-based payments create the risk for landholders of a poor environmental outcome and low payment, which may result in low participation rates. Tendering such incentive contracts further increases conservation effort, but at a decreasing rate as the proportion of performance payments increases. These results highlight the trade-off between maximising the expected level of environmental outcome and maximising budgetary cost-effectiveness.

3.3 Randomised Control Trials

RCTs have been used to evaluate agricultural policy interventions mainly in developing countries. To the best of the authors' knowledge, up to now there exist no applications to the CAP. The examples selected serve as illustration of how the RCT methodology can be applied to a wide variety of policy interventions, including programmes on farmers' training, fertilizer subsidies, agricultural loans, improved seed varieties, and the provision of marketing information. Most of these interventions target specific problems of the agricultural sector at a very local scale, and its evaluation may therefore encompass different stakes and constraints than the evaluation of the EU-wide Common Agricultural Policy. Although the studies reviewed may be only remotely relevant for informing the design of agricultural policy in the EU, the RCT methodology itself promises some potential for that purpose, as will be discussed in more detail in the chapter 5 of this report.

The RCT by Duflo et al. (2011) explores the effectiveness of alternative policy interventions to incentivise Kenyan farmers to use fertilizer. The authors state that farmers in Western Kenya fail to take advantage of apparently profitable fertilizer investments for two possible reasons: (1) unaffordability due to lack of cash or access to credit; or (2) issues of timing and

impatience: at harvest time farmers do have cash available but are not motivated to buy fertilizer in advance. Later in the season when fertilizer is needed there is no cash left. The authors thus hypothesise that policies which encourage farmers to buy fertilizer immediately after harvest could increase fertilizer use. They test whether adaptations to an established subsidy programme, the 'Savings and Fertilizer Initiative (SAFI)', can enhance fertilizer use. More precisely, the experiment tests whether incentives in the form of time-limited discounts on the cost of acquiring fertilizer just after harvest, lead to an increase in farmers' use of fertilizer.

In a first season, a sample of farmers received the SAFI programme in its basic form: the basic SAFI programme provided fertilizer immediately after harvest, when farmers tend to have cash on hand, and offered free delivery, on a date of their choice, of any combination of planting or top dressing fertilizer. Farmers had to make a decision and purchase immediately after harvest.

In the second season, a new sample of farmers was selected, who were randomly assigned to one of the following five groups, each receiving a different version of the SAFI programme: Group 1 received a basic SAFI offer as described above. Group 2 was offered SAFI with the option to freely choose the timing of fertilizer purchase. Group 3 was offered free delivery of fertilizer through SAFI close to the time fertilizer needs to be applied, with payments due at that time. To calibrate the effect of a discount, Group 4 was visited during the same period as Group 3 and offered fertilizer at a 50 percent price discount. This allows comparing the effect of a heavy (50 percent) subsidy offered during the growing season to the effect of the small nudge (in the form of free delivery) offered by the SAFI programme (Group 1). A fifth comparison group who received no incentive to purchase fertilizer served as a reference.

The farmers were randomly selected into each of the treatments. The main outcome of interest was the fertilizer use, with fertilizer purchase through the programme as an intermediate outcome.

The results suggest that an early small discount covering the delivery costs for fertilizer induces the same increase in fertilizer adoption as a heavy subsidy offered later in the season.

Bulte et al.'s (2014) analyse a policy intervention aiming at increasing the use of a modern cowpea seed variety in central Tanzania. Their experimental set-up aims to control for the fact that the yield increase resulting from using improved seed varieties may be partially related to farmers changing their behaviour (for example, providing more effort because of their high expectations on this new variety). In order to do so, the authors compare an 'open' and a 'double-blind' RCT. This consists at the same time of an important methodological innovation, because in policy experiments double-blind interventions are not the standard. The authors expect that, since the participants of the open RCT group know they are 'treated', they will adjust their behaviour. To gauge the potential magnitude of this behavioural effect, the authors contrast such an open RCT to a double-blind trial. 583 farmers were randomly selected to participate in the experiment. . The participants were randomly divided up in four groups. Each group received randomly cowpea seed of either a modern (improved) type or the traditional, local type of cowpea seed. In addition, participants in the open RCT were informed of the type of seed they had been given, while the participants in the double-blind RCT did not know whether they had received seed of an improved or a

traditional variety. Participants in all treatments were informed that the modern seeds yield higher and mature earlier. The seed types were indistinguishable in terms of size and colour. This experimental set-up aimed to identify the impact of these different treatments on the total harvest of cowpeas.

The results can be summarised as follows: The harvest of farmers who were informed that they had received improved seeds was greater. But when farmers were unaware of the type of seed allocated to them, the modern seed type did not outperform the traditional type. Therefore the results strongly suggest that effort matters: farmers who were unsure about the quality of their seed (in the double-blind experiment) and farmers who knew they received the modern seed (in the open RCT) put more efforts (i.e. they planted their seed on larger plots) than farmers who knew they received the traditional seed (control group in the RCT). Thanks to this innovative experimental design, the positive impact on production of the modern seed can be attributed to an increase in farmers' effort rather than to the improved variety itself. This RCT points to the importance of behavioural factors, which traditional policy evaluation techniques would have overlooked.

The RCT study by Giné et al. (2012) examines ways to improve the pay-back of agricultural loans to farmers in Malawi. The study analyses the impacts of improved personal identification (through fingerprinting) on the behaviour of borrowers in a micro-loan scheme. In the treatment group, each farmer had a fingerprint collected as part of the loan application. An explanation was given that the fingerprint would be used to identify them on any future loan applications. The results show that fingerprinting: (1) did not deter farmers from taking out loans; (2) improved loan repayment, particularly for borrowers expected to have the poorest repayment performance; and (3) led farmers to allocate more land and farming inputs to the most profitable crop (paprika) with which loans were repaid.

Using data from a randomised experiment in rural China, Cai et al. (2015) examine the effect of social networks on the take-up of a weather insurance scheme by rice farmers. The experiment provided financial education to a randomly chosen subset of farmers. The authors find that financial education about insurance and its benefits improves take-up by 43% compared to a reference group which received no such education. Large spill-over effects of financial education on adoption by others were also observed: for untreated farmers, the effect of having a friend that did receive the financial education treatment, lead to an important increase in take-up, equivalent to offering a 15% reduction in the insurance premium. Hence, this experiment highlights the importance of network effects among farmers.

Blair et al. (2013) use an RCT to examine the effectiveness of a farmer training programme throughout Armenia. By comparing the treatment group of farmers who received training to a randomly selected reference group, the authors conclude that training did not increase household income or consumption, nor did it affect intermediate outcomes such as adoption of improved agricultural practices or changes in cultivation of crops, suggesting that longer-term impacts are unlikely to materialise.

Nakasone (2013) applies the RCT methodology to study the role of price information on farmers' marketing outcomes in the central highlands of Peru. A group of farmers in randomly selected villages was provided with mobile phones,

through which they received detailed price information for the most relevant crops in six regional markets. The results show that those provided with the phones received higher prices for their products. Information also made farmers more likely to sell their crops. The study also investigates the possibility of information spill-overs by examining marketing outcomes of households who did not receive the mobile phones but lived in villages where others did. Yet, no evidence for such spill-over effects was found in this study. In a similar experiment, Fafchamps and Minten (2012) assess the benefits of an SMS-based agricultural information service in India. The price information is expected to improve farmers' ability to negotiate with buyers and to enable them to arbitrate better across different sales outlet. The weather information should help farmers reduce crop losses due to storms. They conducted an RCT, in which farmers in some (randomly selected) villages were offered free subscription to the SMS information service. In other villages, only some farmers were offered participation, in order to assess whether the information provided to some farmers is transmitted to other non-participating farmers in the same village. The third group is a control group. The results find no statistically significant effect of the treatment on the price received, crop losses or on the likelihood of changing crop varieties and cultivation practices. They do find an effect on where farmers sell their crop: farmers receiving SMS information are less likely to sell at farm-gate and more likely to sell at more distant wholesale markets. They conclude that even though the information provided did not improve outcomes in India, the provision of price information may still be beneficial in other settings, where markets are more disorganised, segmented or too thin to attract a sufficient number of buyers.

While most RCTs reviewed here are evaluations of programmes supporting agriculture in the developing world, these applications do illustrate the potential of RCTs to inform agricultural policies also in countries with modern agriculture and a well-developed agricultural policy.

4 Is there room for economic experiments in the EU agricultural policy evaluation toolbox?

This chapter aims to assess the potential contribution of experimental approaches to the EU agricultural policy evaluation toolbox. We first present the traditional tools used by the European Commission to evaluate the CAP, including case study analyses, simulation approaches and econometric analyses of observational data. Second, we highlight the recent evolutions in the CAP calling for innovations in the evaluation toolbox. Finally, we summarise the complementarities between experimental and the non-experimental approaches.

4.1 The existing policy evaluation toolbox

Evaluation of EU agricultural policy relies on a large diversity of methodologies.

Data sources traditionally used to evaluate agricultural policy are:

- Farm Accountancy Data Network (FADN): data collected annually for a sample of more than 80.000 farms under the supervision of DG AGRI.
- Farm Structure Survey (FSS): data collected by Eurostat across the entire EU farm population.
- Market data from Eurostat, OECD, FAO…
- Administrative data collected by Member States.
- Qualitative information, often focused on a specific topic, collected through surveys, focus group or stakeholders interviews.

To evaluate the impact, efficiency (right use of resources) and effectiveness (achievement of goals) of a policy, these data are used in case-study analyses, to feed and calibrate simulation models, or can be analysed with statistical and econometric methods.

4.1.1 Case studies

Based on a combination of secondary data with qualitative information collected through interviews with key stakeholders, focus group or ad-hoc surveys, case studies can bring further information on a specific question or area. In ex-post evaluation, case studies allow the analyst to develop the answers to the evaluation questions in the light of the experience of a limited number of regions or countries. For ex-ante evaluation, case studies can for example provide valuable information on stakeholders' likely reaction to alternative policy scenarios.

Although case studies might entail analysis of quantitative data, identifying the net causal impact of a policy remains difficult using a case-study approach, as many other factors may have played a role in defining the final outcome, and data on these other factors is usually not available. Also, the interviewed stakeholders may not provide a

representative picture of all those affected, or people's reported reactions may not correspond entirely to their actual behaviour. Moreover, many case studies are needed to draw general conclusions given the limited scope of each case study. Therefore, this method can be very costly when studying the EU wide impact of a policy change. Actually, case studies serve more often as illustration of the general analysis. They are only one element of a larger evaluation approach.

4.1.2 Simulation models

Simulation models of European agricultural production are the main tool for ex-ante evaluation used in DG AGRI. Examples of models for the evaluation of policy change in the EU agricultural sector include:

- CAPRI (Common Agricultural Policy Regionalised Impact Modelling System): It was initially designed to assess the 1992 CAP reform and adjusted since to simulate other CAP changes at the regional (nuts 2) level.
- AGLINK-COSIMO modelling system: one of the most comprehensive partial equilibrium models for global agriculture used in the generation of baseline projections underlying the OECD-FAO Agricultural Outlook.
- MAGNET (Modular Applied GeNeral Equilibrium Tool): global computable general equilibrium model, with a special focus on the agricultural sector.
- IFM-CAP (Individual Farm Model of CAP Analysis): EU-wide mathematical programming model based on individual farms, used to simulate farmers' decision making under alternative policy options and the resulting impacts on agricultural production, farm incomes and the environment (Louhichi et al., 2015).

These simulation models are used to explore important changes in policy options such as quota abolishment, decoupling of direct payments, trade agreements and bioenergy policies. The main advantage of simulation models is that they generally allow analysing the impact of a policy on various criteria such as economic performance (e.g. change in farmers' income) or environmental performance (e.g. nitrogen use).

The main limitation of such models is their reliance on many assumptions and exogenous parameters. For example, in most cases, the underlying decision model is the maximisation of farmers' expected profit under various constraints relating to the availability of fixed variable factors of production, to the structure of technology and to production costs. It can become difficult to discern whether the results obtained from such simulations come from the data or are driven by the technical assumptions. Systematic sensitivity analysis must be undertaken as results of simulations may greatly depend on the choice of some parameters (the calibration issue).

4.1.3 Statistical and econometric analysis of survey data

The econometric analysis for EU agricultural policy evaluation is facilitated by the availability of high-quality agricultural statistics, both at Member State and EU level.

The Farm Accountancy Data Network is a very useful monitoring instrument to evaluate ex-post the impact of CAP changes on farm income, crop choices, input use etc. This large data base is representative of the

EU farming sector in many dimensions for "commercial farms", but so-called small farms are not in the sample, which can induce some selection bias in estimations using FADN data.

The main limitation of the use of observational data (such as FADN data) for policy evaluation is the difficulty to obtain data corresponding to the counterfactual, in order to assess the causal impact of the policy (see 2.2.1). Studies therefore rely on the comparison of one group which is affected by the policy to a group which is not, or they compare the situation before and after the policy. Yet such a between-groups comparison is only valid if the group affected by the policy is perfectly similar to the group which is not affected; in practice this is rarely the case. Selection criteria or self-selection causes the two groups to differ. Also, a before-after comparison, which compares the outcomes after the policy to the baseline situation, is usually not satisfactory since most likely also other factors which affect the outcome variable have changed over the period under study.

More sophisticated econometric methods have been developed in order to identify the causal effect of a policy from observational data (EC 2013b; Loi and Rodrigues 2012). Since many of these methods consist of artificially 'constructing' or mimicking the counterfactual, these methods are also called quasi-experimental approaches. The following empirical strategies aim to construct a proper counterfactual from observational data in order to isolate causal impacts: instrumental variables estimations, regression discontinuity designs, difference-in-difference (DID) matching and propensity score matching (PSM).

For example, Michalek et al. (2015) combine the DID and PSM approach to estimate the extent to which investment aid under the RDP complements or substitutes investments that would be made in the absence of the policy. More precisely, they measure the deadweight loss, i.e. whether the investment support beneficiaries would have undertaken comparable investments even without the investment support. They use data on 1333 dairy farms in Germany for the period 2001-2008. PSM is used to match similar farmers based on their most significant characteristics, and especially on their probability of participating in the programme. The DID approach is then used to estimate the difference in investment behaviour over time between the matched farmers to identify the net causal effect of investment support. Another example of the use of quasi-experimental techniques with an application to the impact of agri-environmental schemes in France is given at the end of this chapter (Illustration 1).

Note that in some occasions, a naturally occurring change may be used for identifying causal effects, without the need to reconstitute a counterfactual. In such 'natural experiments', the researcher takes advantage of a change in context or setting for a sample of the population that has occurred naturally (for example an unexpected abrupt change in legislation). Hence, natural experiments are real situations which resemble to experiments that an experimenter would have conducted with a control group and a test group. However, good natural experiments are the exception. While occasionally such a natural experiment may provide a good opportunity for causal impact evaluation, one cannot 'wait for them to happen'. They should therefore not be relied upon as a general policy evaluation tool.

4.2 Enlarging the current CAP evaluation toolbox

The Common Agricultural Policy (CAP) has experienced several reforms over the last

decades. In order to meet the CAP's evolving objectives, the type of policy interventions have changed drastically. As a result, also the toolbox for evaluating these policies needs to be constantly adapted to follow these policy changes.

Recent evolutions in the CAP call for evaluation methods (1) at the farm or even plot level; (2) accounting for the voluntary nature of many measures; (3) allowing the evaluation of purely regulatory measures (with no EU funding involved); and (4) flexible enough to account for the heterogeneity in CAP implementation across Member States.

First, the targeting of support on farmers rather than commodities and the switch to decoupled payments has called for a change of the unit of evaluation: the farm rather than the market. Simulation models such as CAPRI were initially designed to estimate the market impacts and supply responses to changes in guaranteed prices and other market interventions, and are therefore less suited to analyse impacts at the farm level. In response to this concern, new modelling tools focusing on individual farm behaviour are currently being developed by the European Commission, to capture the farm-specific implementation of policies, as well as the heterogeneity of impact across farmers (Kamel Louhichi et al. 2015). Analysis at the plot level is also required to assess the environmental impacts of certain policies.

Yet, typically the objective function of such farm-level models is often simplistic and limited to profit maximisation. New modelling challenges include the introduction of risk and uncertainty, and accounting for other behavioural factors that may explain heterogeneity in farm decision making. Yet, often information is missing on which behavioural factors are relevant or how they should be modelled. Experiments could help gain insights in developing behavioural utility functions, and to obtain values for important parameters such as risk aversion or time preferences.

Second, an increasing number of CAP measures consist of farmers' voluntary enrolment, such as agri-environmental contracts in the Rural Development Policy. To evaluate the efficiency of such measures, one needs to pay more attention to the drivers of individual farm decisions and on farmers' motivations to participate in such schemes. Ex-ante evaluation should be able to provide information regarding the expected uptake of a voluntary measure, which types of farmers are expected to subscribe, and what incentives are needed to encourage uptake. This type of ex-ante information can help to fine-tune policies so as to obtain the desired level of participation, or to optimise the environmental outcome for a given budget.

Third, the EC is committed to evaluate in a proportionate way all EU spending and non-spending activities intended to have an impact on society or the economy (EC 2015). Some aspects of agricultural policy are strictly regulatory, but were not systematically evaluated in the past. Note that in the regulatory area there is much less data available, compared to the detailed financial data related to spending activities. Data generation methods such as experimental approaches could have potential here as well. Moreover, when farmers receive no payments in exchange for complying with a set of rules (for example with regard to the impact of their activities on the environment), the traditional profit maximising assumption is not useful to understand farmers' reaction to the new regulation, and to predict the compliance rate. Experiments allow capturing different objectives, such as the desire to comply with the rule or to behave like the group (social norms).

Last but not least, the CAP has become more and more diversified in its implementation at the Member State level (both in the first and second pillar). As a result, general EU-wide approaches will increasingly need to be replaced or complemented by more targeted evaluations, focusing at specific policy measures in specific Member States. Case-study approaches have already been mobilised for the evaluation of such local implementations of the Rural Development Policy (RDP). For example, Häring et al. (2004) evaluate the impact of organic aid schemes implemented under the RDP in specific Member States, using a case study approach in 6 Member States. Experimental approaches can complement such qualitative case studies by increasing the internal validity of results obtained from ad-hoc surveys. Moreover, experiments offer a structured evaluation tool, which allows replication in different contexts and the option to adapt them to the specificities of the policy and the location. This largely facilitates the comparison of results across different locations and/or Member States.

4.3 The complementarities between experimental and non-experimental approaches to policy evaluation

Given this needs for innovation in the CAP evaluation toolbox, we argue that economic experiments can interestingly complement non-experimental methods. According to 'Better regulation guidelines' (EC 2015) "Evaluations are based on the best available evidence (…), which should be drawn from a diverse and appropriate range of methods and sources (Triangulation)". This section provides an overview of the important elements to take into account before selecting the best mix of methods for the evaluation of a specific policy.

To shed light on the complementarity of experiments with other methodologies currently in use by policy evaluators, the relative advantages and limitations of different evaluation methods are summarised in Table 3. The left-hand part of the table summarises the non-experimental approaches: "qualitative methods" are mostly based on case studies and qualitative comparisons of before-after policy implementation situations. In "simulation models", we include partial and general equilibrium models, and mathematical programming tools mostly used for ex-ante evaluations. "Statistical & econometric analysis" refers mostly to ex-post evaluations and includes quasi-experimental approaches. Quasi-experiments are sophisticated econometric methods which have been developed to identify the causal effect of policies from observational data (EC 2013b; Loi and Rodrigues 2012) by artificially 'constructing' or mimicking the counterfactual. They include the following empirical strategies: instrumental variables estimations, regression discontinuity designs, difference-in-difference matching and propensity score matching.

The comparison is based on three criteria: the objective of the evaluation; the trade-off between causality and generalizability (i.e. internal and external validity); and the practical challenges in data collection and analysis. The performance indicators (low, medium, high) in Table 3 are just indications of relative performance across methodologies and criteria. They are debatable since each column refers to very heterogeneous evaluation studies. Overall, the assessment of the relative global performance of the different methods depends on the weight attributed to each of the selection criterion.

Finally, we have selected three examples to illustrate the complementarities across

approaches, both between experimental and non-experimental methods (illustration 1), between experimental and quasi-experimental approaches (illustration 2) and between two types of experiments (illustration 3).

4.3.1 Objective of the evaluation

Depending on the main objective of the evaluation task, different methods may be appropriate or may need to be combined. We identify the ability of the different methods (1) to test a policy prior to implementation (ex-ante evaluation); (2) to measure the net impact of a policy (ex-post evaluation) and (3) to understand farmers' reactions to policy in the presence of behavioural factors.

For *ex-ante* evaluations, experimental data and simulation models are highly complementary. Simulation models are clearly more useful when assessing broad policy reforms, such as reducing price support or abolishing production quotas at the EU-level. Yet, when it comes to new policies that are very different from existing ones (e.g. launch of a conservation auction, introduction of novel insurance tools), it may be difficult for simulation models to make realistic assumptions as to the impacts of those policy changes on farmer behaviour. Nevertheless, economic experiments may provide some insights as to what response to expect. For example, the DCE related to the introduction of the greening provisions of the CAP (Schulz et al. 2014) informs policy makers of farmers' likely responses to this new element of the CAP.

When it comes to targeted policy interventions (on specific farm types or in a limited number of regions), simulation models are not the most appropriate method given the cost of making the model flexible enough to capture policy specificities at regional or individual level and because of the lack of data to feed such a model.

Qualitative assessment based on stakeholders' interviews in a number of carefully selected case studies can provide useful insights there, but economic experiments can offer more robust evidence. For example, DCE can help assess how farmers would respond in aggregate to different designs, but also how responses might differ among different types of farmers.

RCT can provide a powerful tool for ex ante analysis in the form of pilot programmes: provided proper randomisation has been set in place, the causal impact of a specific policy programme can be reliably assessed before scaling it up to the entire population. On the contrary, RCT are not suitable to test broad policy reforms (Goldin et al. 2012; Rodrik 2008). For example, in the case of market price interventions it is simply impossible to exclude a (random) part of the population from the policy. RCT can therefore be useful to test specific policy programmes in a specific context, but cannot offer an evaluation of broad or wide-ranging policy reforms.[10]

Also, when behavioural drivers of farmers' responses to policy are important, simulation models may need some input from experimental data. Field experiments can help elicit farmers' preferences and behavioural parameters (e.g. risk aversion, time preferences), which can then be plugged into simulation models. With respect to policy measures with voluntary enrolment, the review of the literature

[10] Opponents of experimentation usually claim that biased answers to big explanatory questions are more important than unbiased answers to smaller casual-descriptive questions (Plott 1989).

above clearly illustrated the usefulness of all types of economic experiments in an ex-ante evaluation perspective.

When considering the *ex-post* evaluation of policy impacts, statistical analysis of observational data and RCT can provide reliable estimates of the impact of a policy or programme. Yet, they usually provide limited information on the reasons underlying the outcomes. When it comes to *understanding why* a policy did not work as expected and how to improve it, qualitative methods based on stakeholders' perceptions, and discrete choice, lab or field experiments are often more useful. Indeed, understanding the impact of a policy requires analysing both the results of the decisions taken by farmers in response to the policy (as can be observed from micro-level data such as FADN), but also to understand the factors underlying such decisions and behaviours. Experiments can help explain unanticipated effects of the policy or intervention (Gneezy and Rustichini 2000). For example, DCE allow analysing in detail the mechanism of farmers' response to a programme, to explain low participation rates in schemes that have failed to attract the envisaged numbers of farmers despite high financial incentives. Hence, choice data will enable the analyst not only to identify the obstacles to participation but also to devise effective remedies.

4.3.2 Trade-off between causality and generalizability

The second key issue in the selection of evaluation methods is the trade-off between internal validity and external validity of the evaluation results. Internal validity or causality reflects the extent to which the causal relation between two variables (for example the policy and the outcome) is properly demonstrated. External validity or generalizability refers both to the ability to transfer the results to the real world and to other contexts. While both internal and external validity are desired, the evaluator is usually confronted with a trade-off (Roe and Just 2009).

Qualitative methods have a low internal validity, since they generally do not allow the analyst to conclude decisively whether the observed changes are due to the policy. Their external validity can be high, if the relevant key stakeholders are interviewed. Nevertheless, stakeholders' stated preferences and future intentions (before a policy is implemented) or perceptions of what has happened (ex-post) can be biased and misreported (strategic bias) or not. Moreover, qualitative methods are often applied in a case study setting, and the transferability of findings to a different context is dependent on the specificities of the case studies.

As long as representative and carefully collected data are used, statistical analysis of observational data presents high external validity. It may be difficult, however, to establish internal validity when the analyst suspects that unobserved variables may have affected outcomes. Michalek et al. (2015) provide an example of how sophisticated, quasi-experimental econometric techniques can be used to estimate the net impact of investment support under the CAP's Rural Development Policy based on FADN data, but such techniques require detailed databases and cannot always provide a convincing solution. On the contrary, random assignment of the participants to the different treatments and high control of the decision-making environment, as used in most economic experiments, allow the researcher to prevent systematic differences in treatment and control groups and to limit any concurrent third elements that could confound the outcome.

However, this strong internal validity comes at the cost of reduced external validity, especially for lab experiments in an abstract setting. Indeed, the choices that individuals make in the lab depend not just on financial implications, but also on the particular context in which a decision is embedded, and the manner in which participants are selected to participate. Because the lab systematically differs from most naturally occurring environments on these dimensions, experiments may not always yield results that are readily generalizable (Levitt and List 2007a). Their abstract setting limits the ability to generalise the results isolated in the lab to real situations where the context is important. Likewise, results obtained with students may not generalise to the targeted population (see box 2). Hence, the ability to transfer conclusions from the lab to the real world may be low.

Box 2: Why running economic experiments with students?

In many experimental studies, the first step consists of running lab experiments with students. At first glance, policy makers may be reluctant to use lab experiments for policy evaluation. But there are several good reasons for that:

First, there are some practical challenges in setting up experiments with stakeholders. Most field experiments require having a fixed number of participants taking part in the experiment simultaneously and in the same room. Of course, this is far easier to organise with students than with busy professionals.

Second, starting with lab experiments is useful because if it does not work as anticipated in the lab, then it is very likely to fail in a field application as well (Plott 1997).

Third, the way in which people behave is, to a certain extent and in certain domains, common across humans. Some findings on the biases and heuristics have been replicated over time and across domains, and can safely be assumed to be valid for any population (van Bavel, Rodríguez-Priego, and Maghiros 2015).

Nevertheless, the lab experiment literature suggests that care should be taken in extrapolating the results obtained from people who have very little experience with the types of decisions being studied (e.g. students) to the real population under study (e.g. famers).

Many studies have shown that: (i) people tend to behave more rationally the higher the financial stakes; (ii) students sometimes behave differently than experienced market participants; (iii) the altruism observed in laboratory experiments sometimes disappears in naturally occurring markets or, on the contrary, more cooperation is observed within the real population under study than students; and (iv) even highly knowledgeable and experienced individuals can act irrationally in laboratory settings because of the artificial way choices are framed (List and Lucking-Reiley, 2002; Alevy et al., 2007; Levitt and List, 2007; Bchir, 2014).

On the other hand, lab experiments are typically very easy to replicate because of the high degree of control over the experiment. This allows testing in how far laboratory results are valid in different contexts (different countries or regions for example).

Field experiments, RCT and DCE present a compromise. They lessen the inherent tension between establishing causality and the generalizability of results. Field experiments give in some control compared to lab experiments, when the experimental

design does not allow controlling for the numerous factors that are at work in the field; but results are likely to be more representative of real-world decision-making. Moreover, even when field experiments are usually conducted on limited samples, they constitute a structured evaluation tool, which allows replication in different contexts or with other participants to test the generalizability of the results. This facilitates comparison of results across different farm types and/or Member States. As such, field experiments provide a bridge between lab and observational data (List, 2007, 2011). A combination of lab experiments with small-scale field experiments can ensure both the correctness and relevance of the results for policy-making.

Since DCEs rely on stated instead of revealed preferences, they may suffer from a lower internal validity than lab experiments. However, because respondents are typically drawn from the targeted population and scenarios proposed are usually as close as possible to what could be observed in reality, external validity may be higher than in an artificial laboratory setting with students. The ability to extrapolate DCE results to other contexts can nevertheless be limited if the survey (e.g. the set and level of attributes) is designed to fit a specific context.

Finally, the RCT methodology combines high internal validity (thanks to randomisation) and high transferability of results to the real world since data come from the natural environment and stakes are real, in contrast to lab and field experiments which are merely games. RCT are therefore often referred to as the "gold standard" to measure the net impact of a specific programme. However, external validity in the sense of transferability of results to a different context or country, or even to a larger scale, may be limited, especially when the policy response is much dependent on specific local factors or when general equilibrium effects play a role (Deaton, 2010).

4.3.3 Practical implementation

The choice of an evaluation method is also often affected by practical issues, due to time and budget constraints. Other criteria might therefore also be important in the choice of an evaluation method, such as the ease of access to data or data collection (cost and time), and the ease of interpretation of evaluation results by policy makers.

Any method requiring the collection of new data will generate additional budget costs and delays. Since the common feature of all economic experiments is that data must be generated under a controlled process, it is expected that they engender greater hassle and data-related costs than simulation models and econometric analysis exploiting existing EU-wide databases such as FADN or Eurostat data.

The time, difficulty and costs involved in collecting observational data or generating experimental data vary greatly. Collection of qualitative data from stakeholders in a case study, or experimental data collected in a discrete choice, lab or field experiment, is usually faster since it involves fairly small sample sizes. For randomised controlled trials, the establishment of the baseline situation and the preparation of the experimental set-up is often a costly and time-consuming precondition for effective experimentation, especially since it may require specific authorization by Member States or by European authorities.

The ease of interpretation of results can also make some evaluation methods more attractive than others. Experimental results are often easier to interpret than the results

from statistical methods or complex simulation models. As indicated by Burtless (1995), “the simplicity of experiments offers notable advantages in making results convincing [...] and understandable to policymakers.” Qualitative data are also easy to interpret since they require limited technical knowledge, although drawing a conclusion from numerous interviews with a diverse pool of stakeholders does not always allow a straightforward interpretation.

Table 3: Comparison of different methods according to the objective of the evaluation and the implementation constraints

	Empirical approaches using observational data			Experimental approaches			
	Qualitative methods	Simulation models	Statistical & Econometric analysis	RCT	Field experiment	Lab experiment	DCE
Main objective of the evaluation	Testing a policy prior to implementation (ex-ante)						
	Medium	High	Low	High (pilot)	High	High	High
	Measuring the net impact of a specific policy (ex-post)						
	Low	Low	Medium (using quasi-experiment)	High	Low	Low	Low
	Understanding farmers' reactions to policy in the presence of behavioural factors						
	Medium	Low	Low	Low	High	High	High
Trade-off between the internal and external validity	Ability to identify causality						
	Low	Not relevant (causal effects are assumed)	Medium	High	Medium	High	Medium
	Ability to transfer the results to the real world						
	Medium (depending on accuracy of stated preference)	High	High	High	Medium (depending on realism of the task)	Low	Medium (depending on accuracy of stated preference)
	Ability to extrapolate results to other contexts/people						
	Low	High (depending on model coverage)	High (depending on data source)	Medium	Medium	Low, but easy to replicate	Medium
Practical implementation	Ease of access to data or data collection						
	Medium	High	High	Low	Medium	High	Medium
	Ease of results' interpretation by policy officers						
	Medium (subjectivity of qualitative data)	Low	Low	High	High	High	High

Source: Inspired by Roe & Just (2009, fig.1 p.1268), extended and adapted for agricultural policy evaluation.

Illustration 1: Predicting EU farms' compliance to greening requirements: Simulation models versus DCE

The policy evaluation question

In the CAP 2014-2020, the greening payment rewards farmers for respecting three mandatory agricultural practices, namely maintenance of permanent grassland, presence of ecological focus area and crop diversification. Farmers' compliance with these requirements is expected to depend upon individual preferences and compliance costs. Before the European Commission formulated the final proposal and before the final vote of the CAP reform was made, the following ex-ante evaluation question was raised: how will EU farmers respond to the greening policy?

We present here the design and results of two studies which allow for the ex-ante assessment of the impact of greening: a simulation model and a discrete choice experiment.

Assessment of different evaluation methods

Simulation model

Louhichi et al (2015) developed an EU-wide Individual Farm Model for CAP Analysis (IFM-CAP) to simulate the effects of the crop diversification requirement on land use and farmers' income. The model is run for every single individual farm of the FADN database, therefore covering all main agricultural production activities and reflecting the heterogeneity of EU farms. Simulation results show that most farms choose to comply with the crop diversification requirement due to the sizable penalty imposed. Yet, the number of farms affected by the crop diversification measure is small and, although some farmers are more affected than others, the overall effect on farm income is very limited.

This simulation approach offers a holistic representation of the farm economic system but, as any simulation model, its results depend on specific assumptions. Farmers are assumed to maximise income and to be risk-neutral, land is assumed to be reallocated only within and not between farms, and several assumptions on supply elasticities, input cost etc. must be made. One of the advantages of such simulation models is that, once the model is calibrated, evaluation results can be obtained and compared rapidly, since it is relatively fast to run simulations. The simulation can also easily be applied at the EU-wide level. In addition, this type of simulation model allows the analyst to test various alternative scenarios and to evaluate the intended effect of a programme at its design stage. Finally, simulation models allow attributing the simulated impact to different underlying mechanisms (at least as far as these mechanisms are considered in the model set-up).

Discrete Choice Experiment

Schulz et al (2014) assess farmers' willingness to comply with the greening requirements using a DCE conducted with 128 German farmers. This study, presented in section 3.1.2, identifies the variables affecting the likelihood to comply with greening requirements and analyses in particular the impact of different designs of the ecological focus area measure.

The main advantage of the DCE over the simulation model approach is that no behavioural assumptions must be made. In the DCE survey, the respondent chooses his/her best option according to his/her

preferences. Both extrinsic (e.g. complying costs, level of green payment) and intrinsic (e.g. personal preferences and social norms) motivations are taken into account. Moreover, the DCE allows exploring the heterogeneity of preferences among farmers. However, a limitation lies in extrapolating the results from a small localised DCE to the EU-wide level. The experiment should be repeated on a larger sample and in different Member States to obtain results valid at the EU level.

In both methods, we don't know how farmers would react with real stakes. Farmers' decisions are simulated based on modellers' assumptions and on calibrated data in the first case and farmers make hypothetical choices which have no real impact in the DCE. Yet, this is a limitation shared by most ex-ante evaluation methods, except if a pilot is conducted, for example through a small-scale RCT.

Illustration 2: Evaluating the impact of agri-environmental schemes (AES): quasi-experiment versus RCT

The policy evaluation question

Agri-environmental schemes are voluntary contracts where farmers receive a payment in return for adopting environmentally friendly farming practices. The advantage of the voluntary approach is that the contracted farmers are supposedly those with the lowest compliance costs, therefore reducing the total cost to reach the environmental objective. However, an AES is effective only if it has an additional effect on farmers' adoption of environmentally friendly practices. Paying for practices that would have been adopted even in the absence of the AES is called a "windfall effect". It increases the cost of the policy and has no additional effect on the environment. Such environmental improvements that would have been made by farmers anyway should be excluded when measuring the net impact of the policy.

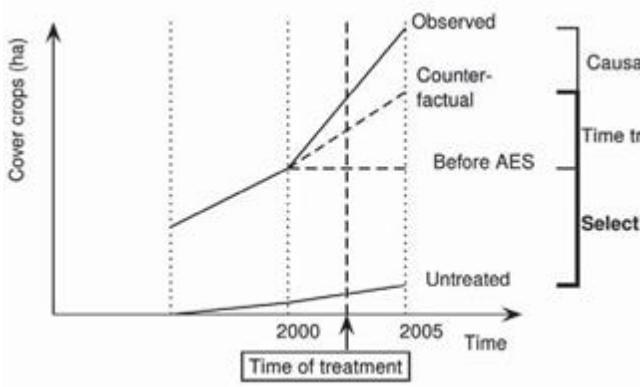
Assessment of different evaluation methods

Quasi-experimental approaches

Chabé-Ferret and Subervie (2013) estimate the net impact of five AES in France, one of which is a subsidy for planting cover crops. The evaluation aims to compare the surface of cover crops grown by farmers having signed an AES contract (the treated) and the surface of cover crops they would have planted in the absence of the AES (the counterfactual). A simple comparison of the situation before the AES (in 2000) and after the AES (in 2005) would be biased given that the surface of cover crops increased over time (Figure 2). Besides, volunteer farmers who contract AES are not similar to those not contracting (see untreated and observed before 2000 in Figure 2). Thus, approximating the counterfactual by the group of farmers who did not engage in AES would lead to an overestimation of the impact of the AES on the planting of cover crops, due to the selection bias.

The authors rely on a quasi-experimental approach, namely difference-in-difference (DID) matching, to estimate the average treatment effect on the treated (ATT). The method relies on matching each engaged farmer (treated) with a farmer not engaged (untreated) but with characteristics similar to those that did engage. While DID-matching is a powerful technique to measure the net impact of a policy, they do require appropriate data and the estimation techniques require some specific assumptions to be fulfilled.

Figure 2: Selection and time trend biases



Source: Chabé-Ferret and Subervie (2013)

Randomised Control Trial

What would be the advantages of a RCT over the use of quasi-experimental techniques based on observational data? Randomisation of beneficiaries would allow computing an unbiased estimate of the net impact of the policy without the need to rely on assumptions as with quasi-experimental techniques. Second, the policy impact (ATT) is obtained through a simple comparison of means, without the need to rely on complex estimation/matching techniques. However, RCT also have limits and raise practical challenges. For instance, it might be politically difficult to offer the AES programme only to a subsample of the population and to refuse enrolment to the rest of the population that would like to enter. To overcome this problem, proposals of "close-to-random" experimental designs are proposed in chapter 5.

Illustration 3: Can contracts with a collective dimension improve AES enrolment and efficiency? DCE versus laboratory experiment

The policy evaluation question

By providing incentives for groups of farmers to enrol in AES (for example by making the financial incentive dependent on a minimum share of farmers enrolled), a dynamic of change can potentially be initiated, and this could increase participation. However, at first sight this may not seem to be the right solution to reduce AES expenditures since it may lead to increased payments to farmers, unless farmers' preferences for such contract result in farmers willing to participate for lower individual payments. Whether introducing a collective dimension in AES contract would increase the effectiveness of the scheme has been tested with two different experimental approaches.

Assessment of different evaluation methods

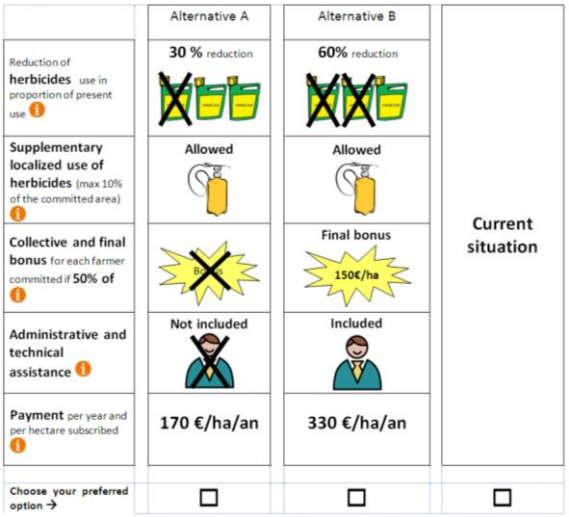
Discrete Choice Experiment

In Kuhfuss et al (2014), farmers' preferences for a bonus conditional on the achievement of a collective enrolment target in an AES were measured using a DCE, conducted on a sample of 317 wine-growers in Languedoc Roussillon (South East of France). Through a web survey, respondents were invited to select their preferred herbicide-reduction contract in different choice sets (Figure 3). One of the attributes of the contract was a bonus of 150€/ha of enrolled land, conditional on 50% of the area of the target territory

being enrolled at the end of the 5-year contract duration. This conditional bonus would be paid in addition to the usual payment at the end of the contract.

The analysis of farmers' stated choices showed that introducing a conditional bonus improved the probability to enrol and, for equivalent environmental effort intensity, it reduced the individual willingness to accept compensation for signing a contract by an amount which, on average, was six times greater than the expected amount of the bonus. Overall, the cost-effectiveness of the AES would therefore be enhanced with a collective bonus.

Figure 3: Example of choice card



Source: Kuhfuss et al (2014)

Laboratory experiment

Le Coent et al (2014) examine a similar type of conditional incentive using a decontextualized laboratory experiment with 220 students in the University of Montpellier (France). This experiment compares the effectiveness and efficiency of two types of incentive mechanisms: contracts with individual payments proportional to individual environmental efforts (similar to existing AES) and contracts with the same individual payments but conditional on the attainment of a collective threshold of environmental efforts. This mechanism presents the obvious advantage of spending public money only if the environmental threshold is reached, but it entails more risks for the contractors of the AES than the contract with a conditional bonus, where the individual basic payment is paid even if the threshold is not reached (Kuhfuss et al., 2014). To avoid biases due to participants' attitudes about environmental issues, the experiment was "context-free": neutral language was used and the individual environmental effort was framed as a financial contribution to a public good. The results of the experiment show that the incentive payments with collective conditionality improve efficiency and effectiveness of AES.

Although relying on two different types of experiments, both studies suggest that conditional financial incentives may interact with individual preferences for collective effort and could contribute to improved enrolment without increasing expenditure.

The main advantage of the lab experiment is the high control over the participants' environment and therefore the high internal validity of the result. It allows concluding that indeed people act differently when a payment is conditional on a collective threshold (although according to theoretical predictions they should not). Because of the decontextualized settings, the lab experiment thus comes to a general observation on human behaviour. However, the lab experiment cannot guarantee that farmers would react similarly to conditional incentives applied to AES.

The external validity of the results of the DCE, which was conducted with farmers with explicit reference to agri-environmental contracts, is expected to be

higher. However, in contrast to participants in the lab experiment who faced real financial stakes, respondents to the DCE were only providing hypothetical answers, since respondents were not paid according to their choices. The different strengths and limits of these two experimental approaches illustrate their complementarity.

5 Practical challenges when implementing economic experiments

When opening the policy evaluation toolbox for experimental approaches, one needs to be aware of a number of limitations, practical issues and constraints related to the implementation of experimental methods. This chapter describes the main practical issues associated with the setting up of experiments: sampling and recruitment of participants (5.1); assignment of participants to the different treatments (5.2); social acceptability and institutional constraints (5.3); other organisational challenges and related costs (5.4 and 5.5);

The objectives of this rather technical discussion on the implementation of experiments are twofold: First, to provide information for evaluators wishing to implement experimental approaches. Second, to make policy officers, who plan to commission evaluation studies, aware of these practical difficulties in order to help them conduct a critical analysis of the proposed methods. For more general advice on conducting behavioural studies, the interested reader can have a look at the JRC report "Seven Points to Remember when Conducting Behavioural Studies in Support of EU Policy-making" (van Bavel, Rodríguez-Priego, and Maghiros 2015).

5.1 Sampling

The first concern of any evaluator is to make sure the group under study (here the participants in an experiment) is representative of the population of interest in order to ensure that results can be generalised to the entire population. EU policy officers will also often aim for results that are valid across Europe rather than being limited to one country (van Bavel, Rodríguez-Priego, and Maghiros 2015).

Note that sample selection is not an additional problem for experiments compared to traditional surveys. The same sample selection problem exists in any survey method. Still, the fact that not all experiments can be run via the internet or at respondents' home may be an extra difficulty compared to traditional surveys.

For laboratory experiments conducted with students, the risk of selection bias is fairly well controlled thanks to the generally large number of participants and to the sophisticated and normalised recruitment procedures followed by most experimental laboratories.

However, this issue is more problematic for field and discrete choice experiments, especially when conducted with farmers. It is very difficult to constitute a sample of farmers who accept to participate in an experiment without departing from the rules of randomisation. Randomisation for experiments run with farmers requires that an up-to-date list of farmers in the region of interest be available to the experimenter who can pick participants at random on the list provided. This is rarely the case since national administrative and statistical services are reluctant to supply this information for privacy protection reasons

and authorization procedures are often slow.

Experimenters therefore tend to resort to other recruitment procedures, for example through calls for participation in magazines or newsletters read by the population under study (for example farmers), or through the network of farm advisors. These methods do not preclude the risk of self-selection. Even if efforts are made to invite farmers to participate on a random basis, one can suspect that their willingness to participate in an experiment is correlated with variables such as their familiarity with administrative or technical staff, their interest in policy-making, their time availability etc. which makes them statistically different from the population of interest and introduces what is known as a "self-selection" bias.

With RCTs, the cases of self-selection bias are less frequent since participation is usually imposed by the experimental design. In order to facilitate randomisation, the unit of randomisation can be moved from the individual level to the institution level – e.g. farm advisory services – or at geographical level – e.g. villages or regions.

Overall, achieving a sample that is representative of all EU farmers in the 28 Member States with the use of experimental methodology is generally not feasible within existing budget constraints. Moreover, the comparisons of results from experiments replicated in several regions or Member States is not straightforward if the selection bias is of different nature in the different contexts. One may therefore reconsider the search for "representativeness" (see 6.1.2).

5.2 Assignment bias

When the experimental approach requires setting up sub-groups receiving different treatments (lab and field experiments, RCTs), particular attention must be paid to the procedure assigning participants to treatment and control groups. To ensure that the average effect of the programme or policy measure under study is properly measured, it is necessary that the treated group and the control group have the same observed and unobserved characteristics before the treatment. Assume for example an RCT conducted to evaluate farm advisory service policies. Farmers are selected to benefit from a tailored individualised technical support programme. The objective is to measure whether such policy can improve the take up of innovations at the farm level. However, if selected farmers are systematically more skilled or are more prone to take risks than farmers of the control group, the probability that they adopt innovations is greater even in the absence of the individualised technical support programme.

Randomisation at the assignment stage is a way to guarantee that the assignment of participants either to treatment groups or to control groups is not correlated to any of their characteristics that could affect the outcomes of the experiment. If random assignment is not correctly done, experiments may overestimate the impact of the policy (Berriet-Solliec et al., 2014).

The ideal procedure to ensure correct randomisation is to use assignment concealment techniques: the investigator in charge of contacting potential participants is not aware of the procedure allocating participants to one group or another and does not take part in the assignment decision procedure (Jaddad and Enkin, 2007).

When it is not always possible to guarantee that assignment is purely random, for example when each treated participant remains free to choose to enrol in the programme or not (e.g. voluntary AES), the experimenter will have to ensure that the

actual take-up of the proposed programme does match the assignment plan, even if this requires enlarging the pool of participants.

If sampling and assignment are not correctly done, the study can yield misleading conclusions: results should not be generalised to the overall population, or measured effects will systematically overestimate or underestimate the true causal effect of the intervention. Being aware of these risks of bias helps interpret the results with caution.

5.3 Potential ethical concerns and close-to-random procedures

In a lab or field experiment, not all participants will receive the same amount of money. In particular, participants who are assigned to the control group may miss the opportunity to make extra money on top of the show-up fee offered to all participants.[11] However, experiments often include "within" procedures in which a participant is successively assigned to a control group and then to a treatment group. In any case, participants are not told the gains of other participants, and their decisions in the experiment remain anonymous (the experimenter cannot link the outcome of an experiment to the name of the participant). These precautions preclude inter-personal comparisons and evaluations of the individuals' performance by the experimenter.

The case of RCT is more controversial because treatment is effective: assignment can create ethical problems since it means that treatment is made available to some (the treated group) and denied to others (the control group) (box 3). Many authors have thus questioned fairness and morality of RCTs (see Baele (2013) for a review). Obviously such ethical concerns are more crucial when RCTs evaluate health treatments (since it directly affects the possibility to be cured) than when they evaluate the impact of agricultural policy measures. Of course, those who are denied access to a specific agricultural support programme may feel disadvantaged. Note however, that this situation also arises under other circumstances, for example due to different administrative procedures in various regions or due to budget constraints which may impose a first-come, first-serve allocation.

However, unequal access to policy may be challenged in court or by European competition policies when the trial creates undue inequity between citizens (for example if it is a subsidy program). This problem can be particularly problematic in the European agricultural sector since the CAP edicts fairly strict rules (especially on first pillar's payments) to avoid competition distortions across European farmers. Indeed, experimental programmes may require notification to obtain approval by the relevant European Commission services. This in turn slows down initiatives to launch experimental pilot programmes to evaluate the likely impact of innovative CAP measures.

Box 3: Some thoughts on experiments and moral issues

In response to those who are tempted to reject experiments on ethical grounds, John List (2011), an eminent professor in experimental economics, gives the following arguments:

[11] The ethical code of conduct of lab and field experimentalists requires that all participants get a fee covering their transport expenses and opportunity cost of time.

"Why it is not fair to only give a fraction of the population a potentially beneficial treatment? First of all, it only considers contemporaneous trade-offs. One could easily argue that it is not fair to future generations to bypass learning opportunities that could make them better off. I am personally glad that earlier generations executed experiments to determine the efficacy of promising drugs so that today my father's heart condition can be treated appropriately. Second, even if one insists on everyone receiving treatment, it remains possible to execute an experiment whereby people receive treatment in waves over time."

Recent works, especially in the development literature, have innovated in many different ways to introduce randomisation into existing programmes with minimal disruption. Duflo et al. (2007), Shadish et al. (2002) and Morawetz (2014) have described "close to random" procedures, that can be used to introduce some randomisation elements in experiments (and therefore reduce the selection bias) without jeopardising acceptance compared to pure randomisation.

- **Randomisation as part of a pilot project:** Small-scale pilot implementation of policy interventions, in which in an initial phase only a small number of people benefit from the policy, after which it will be extended, may face similar ethical concerns as RCT, but seem to be more acceptable. Hence, framing an RCT as a pilot clearly indicates the future extension of the optimised policy, which may contribute to increasing its acceptability. This first close-to-random procedure consists of randomly offering farmers to participate in a pilot study before the measure is implemented at full scale. Those not participating at the beginning serve as the control group; in the end all farmers will be subject to the same policy measure.

- **During the phase-in of a programme**: A similar procedure consists of applying the treatment to participants in successive waves. Financial and administrative constraints can lead to phasing-in of programmes over time, and randomisation can be argued to provide the fairest way of determining the order of phase-in. Randomising the order of phase-in can allow evaluation of programme effects in contexts where it is not acceptable for some groups or individuals to receive no support. The starting point is random and, until all participate, the difference between participants and non-participants can be measured. Note that this design may be problematic when the control group is affected by the expectation of future participation, in which case it does not constitute a valid counterfactual.

- **Over-subscription:** A natural opportunity for introducing randomisation occurs when there are limited resources or implementation capacities and demand for a programme or service exceeds supply. In this case, a natural and fair way to ration resources is to select those who will receive the programme by lottery among eligible candidates.

- **Encouragement design:** Rather than randomising over the treatment itself, participants are randomly assigned to an encouragement to receive the treatment. For example, a random sample of farmers receives by mail an information campaign to participate in a programme. The farms targeted by the campaign are more likely to participate in the programme than others. However, because the encouragement only

increases the probability that a treatment is received without changing it from zero to one, encouragement designs pose specific analytical challenges. Very briefly, having received an invitation to participate is used as a natural instrumental variable with which to evaluate the impact of the treatment (see Duflo et al. (2007) for more details).

- **Free-lunch randomisation** (Morawetz, 2014): From all farms that are eligible to participate in a programme, a lottery selects free-lunch farms. These free-lunch farms are beneficiaries from the programme, irrespectively of whether they comply with the requirements (think for example of an agri-environmental measure) and irrespectively of whether they decided to apply for the measure. This lottery is to be held when the application for participation in the programme is opened but before the programme starts, in order to leave time to inform the free-lunch farms that they do not have to comply with the requirements of the AEM even though they receive the full payment. Such a design allows comparing the outcome for the participant in the programme and the free-lunch farms that were applying, but were drawn in the lottery.

5.4 Strategic biases

Evaluation bias can result from the fact that being part of an evaluation can change behaviour. Such strategic bias arises when participants anticipate that their responses or observed behaviour will have an impact on future policy choices and they wish to manipulate the outcomes of the experiment by adopting an insincere behaviour. Another example of evaluation bias is the 'Hawthorne effect', when the treatment group works harder than normal or the 'John Henry effect' when the control group starts competing with the treatment group. There may also be a "warm glow bias" when the participants are informed on the purpose of the study and change their behaviour in response to their perception of what the evaluator is trying to test.

Each of these biases can affect the reliability of results and may limit the possibility to compare results across studies because such biases are often group or context-dependent.

Controlling for such biases requires the design of the experiment to be carefully done and, when possible, participants not to be made aware of their participation in an experiment. While this is possible for RCTs, it is more difficult for choice, lab and field experiments. Increasing incentives or making the experimental tasks/surveys less prone to manipulation by participants are also a way of limiting this risk.

5.5 Costs and organisational issues

Experiments are often thought to be expensive methods, although this is not necessarily the case. The costs and length of the experiment vary widely from one experiment to another, depending on the scope of the study and the type of experimental design. As a general rule of thumb, laboratory experiments can be expected to be cheaper and yield more rapid results than field experiments and discrete choice experiments, whereas RCTs often display the most costly and time-consuming set-up. It should also be noticed that, while experimental approaches imply costs of data collection, data analysis is relatively quick and therefore cheap, compared to econometric analyses or simulations based on observational data.

Laboratory experiments are a widespread technique (there are more than 70 economic laboratories in the European Union[12]). Most labs have a Charter of Ethics which describes the rules of good scientific practices in experimental research. Amongst others, it is often stipulated, both on ethical and efficiency grounds, that average expected revenue of participating students should be aligned with the opportunity cost of their time. The payment to the participants usually consists of two elements: a fixed show-up fee, corresponding to the minimum guaranteed payoff to all participants, and the payment corresponding to the performance/decisions taken in the game. For European students, it is estimated that it requires an expected gain of 10 to 15 €/ hour plus show-up fees of 2 to 8€. The number of observations necessary to obtain robust estimations depends on the design and research question. On average, a lab experiment is run with 100 to 400 participants and can cost between 2500 and 10000 €.

Switching from the lab to the field can reveal itself more adventurous and more costly for several reasons.

First, participants' gains must be revised upwards to stay in line with farmer's larger opportunity costs of time and to act as an incentive. It may increase the cost of the experiment significantly. In France for example, this suggests that farmers should have average expected earnings of 30 to 50 €/hour.

Second, since it is not easy to bring participants to the lab, especially farmers, experimenters use mobile labs which must reconstitute the conditions of a true lab (including interactions between participants through the computer network, but no oral communication between participants). When it is difficult to gather several participants at the same time and same place, the experimentalist can decide to rely on strategy methods (see box 4). One should therefore add the costs of organisation, including the time and transportation costs of the experimenters, to the actual payments given to participants.

Box 4: Strategy method questionnaire

The use of strategy method questionnaires can make organisation of a field experiment more flexible since they allow conducting experiments even when all the subjects do not play at the same time. This method turns the experiment into an individual survey. In such questionnaires, respondents are asked to specify what would be their decision for each potential strategy that could be chosen by the other players. In other words, they face a lengthy questionnaire instead of playing "in live" with other participants. As in most experiments, responses are incentivised (players can earn a financial reward matching their performance in the game) but if subjects do not play at the same time, then payments must be delayed until the other players have completed their questionnaire.

Strategy method can alter behaviour and is judged to be less acceptable with non-student participants (See Brandts and Charness (2011) for a review). The experimenter thus faces a trade-off: strategy method questionnaires provide a more complete set of responses and simplify the organisational hurdle but responses might be less reliable.

For DCEs, online surveys allow reducing costs compared to DCEs organised in the

[12] Updated figures and list of labs per country are available on the Montpellier experimental lab website :http://leem.lameta.univ-montp1.fr/index.php?page=liste_labos&lang=fr

field. Yet online surveys require that lists of email address of desired participants are available. Online surveys increase the risk of selection bias (most likely only younger farmers would answer) but recent research shows that this risk is declining with the improvement of internet access in rural areas and the increasing familiarity of farmers with new technologies of information, as shown by farm census data indicating the rapidly growing use of the internet by farmers in the EU.

Budget and time necessary to run RCTs are highly variable. They can be fairly low when the RCT consists of sending different information through an existing newsletter to treated groups and non-treated groups to observe their reaction to this information. It can be very expensive if the RCT aims at measuring the impact of a subsidy programme targeting the modernisation of farm holdings since real payments would have to be made to farmers.

Overall, as emphasised by Haynes et al (2012), "with the right academic and policy support, [experiments] can be much cheaper and simpler to put in place than is often supposed. By enabling us to demonstrate just how well a policy is working, [experiments] can save money in the long term - they are a powerful tool to help policymakers and practitioners decide which of several policies is the most cost effective, and also which interventions are not as effective as might have been supposed. It is especially important in times of shrinking public sector budgets to be confident that public money is spent on policies shown to deliver value for money."

6 Conclusion & Recommendations

This report has provided an overview of the usefulness and advantages as well as the limitations and challenges of experimental approaches as new and complementary tools for agricultural policy evaluation. We conclude with a list of six recommendations that result from a careful analysis from the literature, the discussion during the workshop, and the feedback from several DG AGRI policy officers (see Acknowledgments).

6.1 Recommendations to evaluators

The first set of recommendations is directed towards evaluators willing to innovate and introduce experimental approaches in their research in the field of agricultural policy evaluation. The recommendations are formulated both with a view to enhance the use of experimental approaches in CAP evaluations conducted or commissioned by the European Commission, but also for the evaluation of national or local agricultural policies.

1. Make use of the complementarity of methodologies and be open to the use of innovative methods

Although the current tendering process does not exclude experimental approaches, responses to calls for tender are often biased in favour of traditional methods, mostly due to path dependency: the pool of companies and research units responding to calls for tenders in the CAP evaluation domain may implicitly assume that the EC is not open to experimental approaches and may thus refrain from including them in their proposals. Yet, policy officers are fully aware of the need for the toolbox to evolve. In general, calls for tender raise an evaluation question, but leave it to the evaluators to propose the best method. Evaluators can take advantage of this absence of constraints in the methodology in most calls for tenders to propose innovative methods, including experimental approaches.

Furthermore, evaluators should further rely on a combination of approaches, to take the best of all techniques available in the EU evaluation tool box, and benefit from the potential complementarities between experimental and non-experimental methods. For example, behavioural parameters or elasticities estimated through lab or field experiments can be used to improve the parameters of the objective function used in simulation models. DCEs can also provide estimates of the expected adoption rates of rural development programmes, a useful parameter to model their economic and environmental impacts using more complex simulation models.

Good preparatory desk researches, including the use of available statistical data, as well as discussions with stakeholders are necessary steps in the design of experimental studies. Qualitative information is important for the selection of the most relevant type of experiment and treatment variables. Careful analysis of available statistical data is necessary to determine the best sampling plan and sample size of the experiments. Hence, by combining methods, one can exploit the complementarity of approaches.

2. Reconsider the search for "representativeness"

Policy officers generally look for evidence general enough to be valid for the entire population affected by the policy. Given the limitations in sample size and the sampling issues involved in experimental studies (see

5.1), the representativeness and generalizability of the results should be discussed in detail (van Bavel, Rodríguez-Priego, and Maghiros, 2015).

First, given the high level of internal validity of experimental results, their results might still be robust enough to make a solid contribution to the policy-making process even though they may capture behaviour in a specific area only (which may or may not be generalizable to other contexts), especially when experiments are used to complement other methodologies.

Second, many behavioural findings, replicated over time and across domains, can safely be assumed to be valid everywhere and at any time and can therefore help understand reactions to policy of a large share of the EU farming population (e.g. evidence of loss aversion). Some experimental results can be easily extended to a wide range of policy issues and contexts and remobilised in other evaluation processes. For example, experiments to identify efficient designs for agri-environmental auctions provide results on the acceptability and efficiency of different auction formats which are useful in various contexts (Cason & Gangadharan, 2005).

Nevertheless, other behavioural components such as sensitivity to social norms are more local. When the policy change might be affected by behavioural factors which are strongly embedded in cultural characteristics, it is essential to plan for experiments in different Member States. To determine whether a behavioural insight is common to everyone or is specific to a given cultural setting, one can rely on a comparison between a reduced number of countries or regions with different historical and cultural contexts.

Repeating experiments in a number of different contexts may increase the acceptability of experimental results. A careful selection of the case study areas and samples of participants can help ensure that some of the heterogeneity among farmers across EU regions is captured. If no country effects are identified, results could probably be generalised to the EU as a whole (van Bavel, Rodríguez-Priego, and Maghiros, 2015).

3. Proceed by incremental steps: from the laboratory to the field.

Evaluators are encouraged to make use of the complementarities between the different experimental approaches available and rely on a gradual approach: from the lab to the field. This gradual approach in the use of economic experiments can increase the cost efficiency of an evaluation proposal and improve the quality of conclusions.

For example, laboratory experiments are typically appropriate methods to pre-test individual and group responses to different incentives or policy designs. If the policy does not work as anticipated in the lab, it is very likely to fail in a field application as well (Plott, 1997). Moreover, different treatments of a lab experiment allow comparing several initial versions of a policy easily. The experiment plays the equivalent role of wind-tunnel testing for a newly designed aircraft. Then, as emphasised by Hellerstein et al. (2015) from USDA's Economic Research Service, "the next logical step after laboratory testing is a field test - the equivalent of a test flight".

Lab experiments, initially run with students, can then be repeated with farmers, in a more contextualised setting[13]; in order to check the

[13] Running a context-free lab experiment with farmers can be difficult. The strict respect of the instructions and the acceptance of the incentive payment rule are usually harder to obtain than with students. One solution can be to move directly from a context-free lab experiment with students, to a contextualized field experiment with farmers, with words referring to their environment, such as their farming practices. This solution helps to improve the acceptability of the game and its external validity.

sensitivity of outcomes to the context in which the policy is applied. In a final step, RCTs can be envisaged to confirm results, fine-tune implementation and engage a policy dialogue with stakeholders and policymakers.

4. Find responses to moral and ethical obstacles associated to randomisation

All aspects of EU action, including evaluation procedures, require that there is no discrimination between stakeholders. Therefore, when making use of experiments, one should make sure to respect this principle. As long as experiments do not involve effective policy interventions (as is the case for most choice, lab and field experiments), this is not a problem. Yet, the core element of RCTs is the effective implementation of (some aspects of) policy measures to only one, randomly selected, part of the target population.

For any experimental research to be acceptable within the EU context, a detailed description of the approach chosen for addressing this issue of non-discrimination should therefore be part of the proposal. Innovative approaches to randomisation and close-to-random procedures, discussed in detail in section 5.3 may provide a solution.

6.2 Recommendations to policy officers at the European Commission

We conclude by providing a few institutional suggestions that could help develop the use and the efficiency of experimental approaches for CAP evaluation at the European Commission level.

1. Invest in resources and expertise on the use of experimental methods for policy evaluation

With respect to the development of policy proposals and their ex-ante impact evaluation, investing in internal expertise on the use of experimental methods would allow spreading their use to inform policies at the design phase. One example is the newly created 'Unit for Scientific Foresight and Behavioural Insight' of the Joint Research Centre of the EC.

In 2010, the European Commission set up a *Framework Contract for the Provision of Behavioural Studies* (see box 4*).* Its purpose is to facilitate the running of behavioural studies by all DGs in support of EU policy-making. DG AGRI could make use of this framework contract for experimental studies concerning CAP evaluation.

Box 4: Framework contract for the provision of behavioural studies

In January 2012 the European Commission put in place a Framework Contract for the provision of behavioural studies, open to all services. The Framework Contract considerably reduces the administrative workload and the timeframe for undertaking specific behavioural studies that may support the EC policy agenda. The first version of the Framework Contract – which will be running until December 2015 – has already been mobilised by 10 different DGs which, all together, contracted 19 behavioural studies, focussing on diverse issues: from energy labelling to food labelling, from online gambling to the use of credit cards for purchases, to name but a few. From January 2016, the current Framework Contract will be replaced by a new and improved version.

2. Inform evaluators on the potential of experimental approaches and openness towards the use of these new methods

Informing external contractors on the advantages and limitations of experimental approaches for policy-making could help raise awareness of the potential of these new methods and could demonstrate DG AGRI's openness towards the use of these methods in future evaluation contracts. Also, Member States could be informed of the potential of experimental approaches, for example with respect to the evaluation of nationally implemented rural development policies.

Moreover, informing evaluators of the procedure for authorisation of pilot tests and randomisation by the EC services would allow evaluators to take into account the constraints associated to these procedures and may encourage the setting up of RCTs.

References

Alevy, J. E., M. S. Haigh, and J. A. List. 2007. "Information Cascades: Evidence from a Field." *Journal of Finance* 62 (1): 151–81.

Alfnes, F., A. G. Guttormsen, G. Steine, and K. Kolstad. 2006. "Consumers' Willingness to Pay for the Color of Salmon: A Choice Experiment with Real Economic Incentives." *Americal Journal of Agricultural Economics* 88 (4): 1050–61.

Baele, S. J. 2013. "The Ethics of New Development Economics: Is the Experimental Approach to Development Economics Morally Wrong?" *The Journal of Philosophical Economics* VII (1).

Bahrs, E., S. Kroll, and M. Sutter. 2008. "Trading Agricultural Payment Entitlements: An Experimental Investigation of Bilateral Negotiations." *American Journal of Agricultural Economics* 90 (5): 1201–7.

Banerjee, S., F. P. De Vries, N. Hanley, and D. Van Soest. 2014. "The Impact of Information Provision on Agglomeration Bonus Performance: An Experimental Study on Local Networks." *Stirling Economics Discussion Paper* 96 (4): 1009–29.

Bastian, C. T., D. J. Menkhaus, A. M. Nagler, and N. S. Ballenger. 2008. "Ex Ante Evaluation of Alternative Agricultural Policies in Laboratory Posted Bid Markets." *American Journal of Agricultural Economics* 90(5): 1208–1215.

Bchir, M. A. 2014. "From the Lab to the Field: An Experimental Investigation of the Provision of a Club Good." *Journal of Behavioral and Experimental Economics* 51 (August): 57–67.

Beharry-Borg, N., J. Smart, M. Termansen, and K. Hubacek. 2013. "Evaluating Farmers' Likely Participation in a Payment Programme for Water Quality Protection in the UK Uplands." *Regional Environmental Change* 13 (3): 633–47.

Berk, R.A., R.F. Boruch, D.L. Chambers, P.H. Rossi, and A.D. Whit. 1985. "Social Policy Experimentation: A Position Paper." *Evaluation Review* 9 (4): 387–429.

Berriet-Solliec, M., P. Labarthe, and C. Laurent. 2014. "Goals of Evaluation and Types of Evidence." *Evaluation* 20: 195–2013.

Birol, E., M. Smale, and A. Gyovai. 2006. "Using a Choice Experiment to Estimate Farmers' Valuation of Agrobiodiversity on Hungarian Small Farms." *Environmental and Resource Economics* 34 (4): 439–69.

Birol, E. and P. Koundouri (2008): *Choice Experiments Informing Environmental Policy – A European Perspective.* Edward Elgar, 368 pp.

Blair, R., K. Fortson, J. Lee, and A. Rangarajan. 2013. "Should Foreign Aid Fund Agricultural Training? Evidence from Armenia August 2013." *Mathematica Policy Research Working Paper* 19 (August).

Bocquého, G., F. Jacquet, and A. Reynaud. 2013. "Reversal and Magnitude Effects in Long-Term Time Preferences: Results from a Field Experiment." *Economic Letters* 120 (1): 108-111.

Brandts, J., and G. Charness. 2011. "The Strategy versus the Direct-Response Method: A First Survey of Experimental Comparisons." *Experimental Economics* 14 (3): 375–98.

Breckon, J. 2015. *Better Public Services through Experimental Governments*. http://www.alliance4usefulevidence.org/assets/Final.pdf

Breustedt, G., N. Schulz, and U. Latacz-Lohmann. 2013 b. "Calibrating Agri-Environmental Schemes Using a Two-Stage Discrete Choice Experiments." *German Journal of Agricultural Economics* 62 (4): 259–75.

Breustedt, G., N. Shulz, and U. Latacz-Lohmann. 2013 a. "Factors Affecting Participation and Compensation Requirements in Agri-Environmental Schemes: Insights from a Discrete Choice Experiment." *German Journal of Agricultural Economics* 62 (4): 244–58.

Broch, S. W. and S. E. Vedel. 2012. "Using Choice Experiments to Investigate the Policy Relevance of Heterogeneity in Farmer Agri-Environmental Contract Preferences." *Environmental and Resource Economics* 51 (4): 561–81.

Bulte, E., G. Beekman, S. Di Falco, J. Hella, and P. Lei. 2014. "Behavioral Responses and the Impact of New Agricultural Technologies: Evidence Froma Double-Blind Field Experiment in Tanzania." *American Journal of Agricultural Economics* 96 (3): 813–30.

Burtless, G. 1995. "The Case for Randomized Field Trials in Economic and Policy Research." *Journal of Economic Perspectives* 9 (2): 63–84.

Cai, J., A. de Janvry, and E. Sadoulet. 2015. "Social Networks and the Decision to Insure." *Americal Journal of Agricultural Economics* 7 (2): 81–108.

Cason, T. N., and L. Gangadharan. 2005. "A Laboratory Comparison of Uniform and Discriminative Price Auctions for Reducing Non-Point Source Pollution." *Land Economics* 81 (1): 51–70.

Cason, T. N., L. Gangadharan, and C. Duke. 2003. "A Laboratory Study of Auctions for Reducing Non-Point Source Pollution." *Journal of Environmental Economics and Management* 46 (3): 446–71.

Chabé-Ferret, S., and J. Subervie. 2013. "How Much Green for the Buck? Estimating Additional and Windfall Effects of French Agro-Environmental Schemes by DID-Matching." *Journal of Environmental Economics and Management* 65 (1): 12–27.

Choi, J. J., D. Laibson, B. Madrian, and A. Metrick. 2003. "Optimal Defaults." *American Economic Review* 93(2): 180-185.

Christensen, T., A. B. Pedersen, H. O. Nielsen, M. R. Mørkbak, B. Hasler, and S. Denver. 2011. "Determinants of Farmers' Willingness to Participate in Subsidy Schemes for Pesticide-Free Buffer Zones-A Choice Experiment Study." *Ecological Economics* 70 (8): 1558–1564.

Ciriolo, E. 2011. "Behavioural Economics in the European Commission: Past, Present and Future." *Oxera*.

Cook, Thomas D. 2003. "Why Have Educational Evaluators Chosen Not to Do Randomized Experiments?" *The Annals of the American Academy of Political and Social Science* 589 (1): 114–49.

Cummings, R. G., C. A. Holt, and S. K. Laury. 2004. "Using Laboratory Experiments for Policy Making: An Example of the Georgia Irrigation Reduction Auction." *Journal of Policy Analysis and Management* 23 (2): 341-363.

Deaton, A. *2010*. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48 (2): *424-55*.

Duflo, E., R. Glennester, and M Kremer. 2007. "Using Randomization in Development Economics Research: A Toolkit." *Handbook of Development Economics* 4: 3895–3962.

Duflo, E., M. Kremer, and J. Robinson. 2011. "Nudging Farmers to Utilize Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101 (October): 2350–90.

EC. 2013a. *Design and Commissioning of Impact Evaluations. Authors: Morris S., Tödtling-Schönhofer H., Wiseman M., Commissioned by DG Employment, Social Affairs and Inclusion, Unit A3.* http://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=7646.

———. 2013b. “Designing and Commissioning of Counterfactual Impact Evaluations.” http://ec.europa.eu/social/BlobServlet?docId=10805&langId=en.

———. 2015. *Better Regulation Guidelines*. http://ec.europa.eu/smart-regulation/guidelines/toc_tool_en.htm

Espinosa-Goded, M., J. Barreiro-Hurlé, and E. Ruto. 2010. “What Do Farmers Want from Agri-Environmental Scheme Design? A Choice Experiment Approach.” *Journal of Agricultural Economics* 61 (2): 259–73.

Fafchamps, M., and B. Minten. 2012. “Impact of SMS-Based Agricultural Information on Indian Farmers.” *World Bank Economic Review* 26 (3): 383–414.

Giné, X., J. Goldberg, and D. Yang. 2012. “Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi.” *The American Economic Review* 102 (6): 2923–54.

Ginon, E., C. Chabanet, P. Combris, and S. Issanchou. 2014. “Are Decisions in a Real Choice Experiment Consistent with Reservation Prices Elicited with BDM ‘auction’? The Case of French Baguettes.” *Food Quality and Preference* 31: 173–80.

Goldin, I., F. Halsey Rogers, and N. Stern. 2012. “Making Aid Work.” *Boston Review*. http://bostonreview.net/goldin-rogers-stern-economy-whole.

Greiner, R. 2015. “Factors Influencing Farmers’ Participation in Contractual Biodiversity Conservation: A Choice Experiment with Northern Australian Pastoralists.” *Australian Journal of Agricultural and Resource Economics* 58: 1-21.

Greiner, R., and D. Gregg. 2011. “Farmers’ Intrinsic Motivations, Barriers to the Adoption of Conservation Practices and Effectiveness of Policy Instruments: Empirical Evidence from Northern Australia.” *Land Use Policy* 28 (1): 257–65.

Grolleau, G., N. Mzoughi, and S. Thoyer. 2015. “Les Incitations Monétaires Dans La Politique Agro-Environmentale: Peut-on Faire Mieux Avec Moins?” *Review of Agricultural and Environmental Studies* 96 (2): 241–57.

Hanley, N., D. Mac Millan, R. E Wright, C. Bullock, I. Simpson, D. Parsisson, and B. Crabtree. 1998. “Contingent Valuation versus Choice Experiments: Estimating the Benefits of Environmentally Sensitive Areas in Scotland.” *Journal of Agricultural Economics* 49 (1): 1–15.

Häring, A.M., N. Lampkin, J. Tuson, S. Dabbert, J. Aurbacher, B. Bichler, C. Eichert, et al. 2004. *Organic Farming and Measures of European Agricultural Policy*. *Organic Farming in Europe: Economics and Policy*. Vol. 11.

Harrison, G.W., and J.A. List. 2004. “Field Experiments.” *Journal of Economic Literature* 42 (4): 1009–55.

Harrison, G.W. 2011. “Experimental Methods and the Welfare Evaluation of Policy Lotteries.” *European Review of Agricultural Economics* 38 (3): 335–60.

Harrison, G.W., Lau, M.I., Rutström, E.E.(2011) Theory, experimental design and econometrics are complementary (and so are lab and field experiments), In: *The Methods of Modern Experimental Economics*, Eds: Frechette, G. and Schotter, A., Oxford University Press, New York.

Haynes, L., O. Service, B. Goldacre, and D. Torgerson. 2012. "Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials." https://www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials

Hazell, P, and R. Norton. 1986. *Mathematical Programming for Economic Analysis in Agriculture*. New York: MacMillan Publishing Company.

Hellerstein, D., N. Higgins, and M. Roberts. 2015. "Options for Improving Conservation Programs: Insights from Auction Theory and Economic Experiments." *U.S. Department of Agriculture, Economic Research Service,* ERR-181 (January). http://www.ers.usda.gov/media/1732610/err181.pdf.

Herberich, D. H., S. D. Levitt, and J.A. List. 2009. "Can Field Experiments Return Agricultural Economics to the Glory Days?" *American Journal of Agricultural Economics* 91 (5): 1259–65.

Jaddad, A. R., and M. Enkin. 2007. *Randomized Controlled Trials: Questions, Answers, and Musings.* Second Edi. Oxford, UK: Blackwell Publishing Ltd.

Kahneman, D. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review* 93 (5): 1449-1475.

Kawasaki, K., T. Fujie, K. Koito, N. Inoue, and H. Sasaki. 2012. "Conservation Auctions and Compliance: Theory and Evidence from Laboratory Experiments." *Environmental and Resource Economics* 52 (2): 157–79.

Kuhfuss, L., R. Preget, and S. Thoyer. 2014. "Individual Preferences and Collective Incentives: What Design for Agri-Environmental Contracts? The Case of Wine-Growers' Herbicide Use Reduction." *Review of Agricultural and Environmental Studies* 95 (1): 111–43.

Latacz-Lohmann, U., S. Schilizzi, and G. Breustedt. 2012. "Auctioning Outcome-Based Conservation Contracts". *Schriften Der Gesellschaft Für Wirtschafts- Und Sozialwissenschaften Des Landbaus e.V*, Bd. 47: 345–57.

Latacz-Lohmann, U. and C. van der Hamsvoort. 1997. "Auctioning Conservation Contracts: A Theoretical Analysis and an Application." *American Journal of Agricultural Economics* 79 (2): 407–18.

Le Coent, P., R. Préget, and S. Thoyer. 2014. "Designing Agri-Environmental Contracts: Can Collective Conditionality Improve Participation?" *Working Paper*. www.sfer.asso.fr/content/.../jrss2013_a2_lecoent.pdf

Levitt, S. D., and J. A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Tell Us About the Real World?" *Journal of Economic Perspectives* 21 (2): 153–74.

Li, C.-Z., J. Kuuluvainen, E. Pouta, M. Rekola, and O. Tahvonen. 2004. "Using Choice Experiments to Value the Natura 2000 Nature Conservation Programs in Finland." *Environmental and Resource Economics* 29 (3): 361–74.

Lips, M., and C. Gazzarin. 2008. "What Are the Preferences of Dairy Farmers Regarding Their Work? A Discrete Choice Experiment in the Eastern Part of Switzerland." *Paper presented at EAAE 2008 International Congress*, August 26-29, Ghent, Belgium, 6 p. http://purl.umn.edu/44132.

List, J. A. 2007. "Field Experiments: A Bridge between Lab and Naturally Occurring Data." *The B.E. Journal of Economic Analysis & Policy* 5 (2): 1–47.

———. 2011. "Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off." *Journal of Economic Perspectives* 25 (3): 3–16.

List, J. A., and D. Lucking-Reiley. 2002. "The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign." *Journal of Political Economy* 110 (1): 215–33.

Loi, M., and M. Rodrigues. 2012. "A Note on the Impact Evaluation of Public Policies: The Counterfactual Analysis." *JRC Scientific and Policy Report*. http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/26891/1/lbna25519enn.pdf.

Louhichi, K., P. Ciaian, M. Espinosa, L. Colen, A. Perni, and S. Gomez y Paloma. 2015. *An EU-Wide Individual Farm Model for Common Agricultural Policy Analysis (IFM-CAP)*.

Lusk, J. L., and T. C. Schroeder. 2004. "Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks." *American Journal of Agricultural Economics* 86 (2): 467–82.

Maart-Noelck, S. C., O. Musshoff, and M. Maack. 2013. "The Impact of Price Floors on Farmland Investments: A Real Options Based Experimental Analysis." *Applied Economics* 45 (35): 4872–82.

McIntosh, C. R., J.F. Shogran, and E. Dohlman. 2007. "Supply Response to Countercyclical Payments and Base Acre Updating Uncertainty: An Experimental Study." *American Journal of Agricultural Economics* 89(4): 1046–1057.

Michalek, J., P. Ciaian, and D. Kancs. 2015. "Investment Crowding Out: Firm-Level Evidence from Northern Germany." *Regional Studies*, 1–16.

Michaud, C., D. Llerena, and I. Joly. 2012. "Willingness to Pay for Environmental Attributes of Non-Food Agricultural Products: A Real Choice Experiment." *European Review of Agricultural Economics* 40 (2): 313–29.

Morawetz, U B. 2014. "A Concept for a Randomized Evaluation of Agri-Environment Measures." In *The Common Agricultural Policy in the 21st Century*, edited by E. Schmid and S. Vogel, 113–30. Vienna.

Nagler, A. M., D. J. Menkhaus, C. T. Bastian, N. S. Ballenger, E. O'Donoghue, and C. E. Young. 2009. "Are Production Decisions Decoupled under a Bond Scheme? Experimental Evidence." *Review of Agricultural Economics* 31 (2): 222–30.

Nagler, A. M., D. J. Menkhaus, C. T. Bastian, M. D. Ehmke, and K. T. Coatney. 2013. "Subsidy Incidence in Factor Markets: An Experimental Approach." *Journal of Agricultural and Applied Economics* 45 (1): 17–33.

Nakasone, E. 2013. "The Role of Price Information in Agricultural Markets : Experimental Evidence from Rural Peru The Role of Price Information in Agricultural Markets : Experimental Evidence from Rural Peru." In *Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting*.

Normann, H.-T., and R. Ricciuti. 2009. "Laboratory Experiments for Economic Policy Making." *Journal of Economic Surveys* 23 (3): 407–32.

Parkhurst, G. M., and J. F. Shogren. 2007. "Spatial Incentives to Coordinate Contiguous Habitat." *Ecological Economics* 64 (2): 344–58.

———. 2008. "Smart Subsidies for Conservation." *Americal Journal of Agricultural Economics* 90 (4): 192–200.

Parkhurst, G. M., J. F. Shogren, C. Bastian, P. Kivi, J. Donner, and R. B. W. Smith. 2002. "Agglomeration Bonus: An Incentive Mechanism to Reunite Fragmented Habitat for Biodiversity Conservation." *Ecological Economics* 41 (2): 305–28.

Paulrud, S., and T. Laitila. 2010. "Farmers' Attitudes about Growing Energy Crops: A Choice Experiment Approach." *Biomass and Bioenergy* 34 (12): 1770–79.

Plott, C. R. 1989. "An updated review of industrial organization: applications of experimental methods" *Science* II (1982).

Plott, C. R. 1997. "Laboratory Experimental Testbeds: Application to the PCS Auction." *Journal of Economics and Management Strategy* 6 (3): 605–38.

Reeson, A. F., and J. G. Tisdell. 2010. "The Market Instinct: The Demise of Social Preferences for Self-Interest." *Environmental and Resource Economics* 47 (3): 439–53.

Reeson, A., and K. Nolles. 2009. "Experimental economics: applications to environmental policy. CSIRO Working Paper Series 2009-03." *Experimental Economics*, March.

Rodrik, D. 2008. The New Development Economics : We Shall Experiment, but How Shall We Learn ? *Faculty Research Working Papers Series.* John F. Kennedy School of Government, Harvard University

Roe, B. E., and D. R. Just. 2009. "Internal and External Validity in Economics Research: Tradeoffs between Experiments, Field Experiments, Natural Experiments, and Field Data." *American Journal of Agricultural Economics* 91 (5): 1266–71.

Rolfe, J., J. Windle, and J. McCosker. 2009. "Testing and Implementing the Use of Multiple Bidding Rounds in Conservation Auctions: A Case Study Application." *Canadian Journal of Agricultural Economics* 57 (3): 287–303.

Ruto, E., and G. Garrod. 2009. "Investigating Farmers' Preferences for the Design of Agri-Environment Schemes: A Choice Experiment Approach." *Journal of Environmental Planning and Management* 52 (5): 631–47.

Samuelson, P., and W. Nordhaus. 1985. *Principles of Economics*. New Delhi: MacGraw Hill.

Schilizzi, S, and U. Latacz-Lohmann. 2013. "Conservation Tenders: Linking Theory and Experiments for Policy Assessment." *Australian Journal of Agricultural and Resource Economics* 57 (1) 15–37.

Schulz, N., G. Breustedt, and U. Latacz-Lohmann. 2014. "Assessing Farmers' Willingness to Accept 'Greening': Insights from a Discrete Choice Experiment in Germany." *Journal of Agricultural Economics* 65 (1): 26–48.

Shadish, W. R, T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Ho.

Thaler, R. H., and C. R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth and Happiness*. New Haven and London: Yale University Press.

Van Bavel, R., N. Rodríguez-Priego, and I. Maghiros. 2015. "Seven Points to Remember When Conducting Behavioural Studies in Support of EU Policy-Making." *JRC Scientific and Policy Report*. Seville.

Annex 1: Workshop agenda

How can economic experiments inform EU agricultural policy? Considerations for CAP 2014-2020 evaluation

Date: 8 January 2015 14h-18h

Place: DG AGRI AP L130 11 S.1

14h-14h20: Introduction (Marianne Lefebvre)

14h20-14h40: What are the drivers of farmers' responses to agricultural policy? (Sophie Thoyer)

14h40-16h: Presentation of different experimental approaches and examples of economic experiments relevant for CAP evaluation (Uwe Latacz-Lohmann)

16h20-17h05: Value added and limitations of experimental approaches compared to other methodologies currently in use in designing and evaluating the CAP (Raphaële Préget)

Complementarities between simulation models and experimental approaches for policy design (Discussion by Koen Dillen)

The limitations in current CAP evaluation toolbox (Discussion by Anette Hurrelmann)

17h05-17h20: Running experiments for DG SANCO: sharing experiences (Emanuele Ciriolo)

17h20-17h50: Practical limitations and recommendations to facilitate experimentation in the CAP (Sophie Thoyer)

17h50-18h: Conclusion (Sergio Gomez y Paloma)

Annex 2: List of participants

Organizers

Liesbeth Colen	DG JRC – IPTS, Agrilife Unit
Sergio Gomez y Paloma	DG JRC – IPTS, Agrilife Unit
Marianne Lefebvre	DG JRC – IPTS, Agrilife Unit (currently Assistant Professor at Université d'Angers, GRANEM UMR, Angers, France)

Experts

Uwe Latacz-Lohmann	Professor of Agricultural Economics. Kiel University, Department of Agricultural Economics, D-24118 Kiel, Germany, and School of Agricultural and Resource Economics, University of Western Australia, Australia
Raphaële Préget	Researcher at INRA, UMR LAMETA, Montpellier, France
Sophie Thoyer	Professor of Environmental and Agricultural Economics. Montpellier Supagro, UMR LAMETA, Montpellier, France

Policy officers (DG AGRI)

Bruno Buffaria (AGRI C3),	Andreas Kolodziejak (AGRI E4),
Emanuele Ciriolo (JRC E2),	Maciej Krzysztofowicz (AGRI C2),
Flavio Coturni (AGRI E1),	Notis Lebessis (AGRI E01 Advisor),
Gregorio Dávila Díaz (AGRI H1),	Pierluigi Londero (AGRI E2),
Olivier Diana (AGRI H4),	Koen Mondelaers (AGRI E2),
Koen Dillen (AGRI E2),	Yves Plees (AGRI E4),
Tomas Garcia Azcarate (AGRI C1),	Benjamin Van Doorslaer (AGRI E2),
Giampero Genovese (AGRI E2),	Gesa Wesseler (AGRI E3)
Annette Hurrelmann (AGRI B2),	

European Commission

EUR 27496 EN – Joint Research Centre – Institute for Prospective Technological Studies

Title: **(How) can economic experiments inform EU agricultural policy?**

Author(s): Liesbeth Colen, Sergio Gomez y Paloma, Uwe Latacz-Lohmann, Marianne Lefebvre, Raphaële Préget & Sophie Thoyer

Luxembourg: Publications Office of the European Union

2015 – 74 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online)

ISBN 978-92-79-52251-2 (PDF)

doi:10.2791/17634

LF-NA-27496-EN-N

JRC Mission

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

Serving society
Stimulating innovation
Supporting legislation

Publications Office

doi:10.2791/17634
ISBN 978-92-79-52251-2