The Potential of Different Experimental Designs for Policy Impact Assessment

Sven Grüner and Norbert Hirschauer
Martin-Luther-Universität Halle-Wittenberg

Oliver Mußhoff
Georg-August-Universität Göttingen

Abstract

Economic experiments have traditionally been conducted in laboratory settings. Since experimental conditions can be easily controlled and manipulated in the lab, high internal validity can be achieved. The external validity of lab experiments, however, is often poor due to the highly stylized environment. Hence, in recent years, researchers have increasingly left the lab and used the Internet to run economic experiments. In this paper, we aim to systematize economic experiments and discuss the advantages and disadvantages of online approaches. In particular, we focus on the question of how experiments can be used for policy analysis in the agricultural sector. Our core findings are as follows: first, the costs of online experiments are considerably lower than those of traditional lab experiments. This applies to the direct costs of experimenters as well as to the opportunity costs of experimental subjects. Second, experimenters, who always struggle with limited budgets, can exploit the cost advantage of online approaches and take various measures to increase external validity. Spare funds can be used to recruit more participants and/or to grant higher performance-related payoffs. In conjunction with participants' reduced opportunity costs, they will also make it easier to recruit representatives of the social group of interest (e.g., farmers), instead of using convenience groups of students as surrogate experimental subjects. A high-numbered experimental testing of the real behavior of real decision makers who face relevant real payoffs has a good chance to increase the quality of conditional behavioral forecasting. This, in turn, is the prerequisite of reliable policy analysis. Third, decisions in online experiments are made in the familiar setting of people's home offices. The situational context is thus much more similar to decision making in regular life than a lab setting. While being beneficial for external validity, using the home setting also entails a disadvantage. It reduces internal validity because the extra-laboratory decision environment cannot be well controlled. Experimenters cannot observe, for example, which sources of information, tools, time, and effort participants use to arrive at experimental decisions.

1 Introduction

The core idea of a behavioral experiment is to ceteris paribus manipulate an independent variable in an artificially designed and well-controlled setting, thus facilitating the identification of the variable's influence on decision making. Contrary to well-controllable experimental settings, many variables simultaneously and continuously change in real-life environments that are the traditional object of scientific observation. The violation of the ceteris paribus condition in observational studies hampers or even impedes the identification of cause and effect relationships. This especially applies if, in the real-life environment, strong stochastic influences interfere with systematic developments that have only long-term effects.

The potential field of application for behavioral experiments is vast. It encompasses positive analysis (generation of hypotheses, theory testing), as well as conditional forecasting and normative analysis for social actors who attempt to steer the behavior of others through various mechanisms (mechanism design). In the past, theory and hypothesis testing has been one of the most important reasons to run behavioral experiments. A prominent example is expected utility theory that is based on simple rationality axioms (cf., VON NEUMANN and MORGENSTERN, 1944). Its explanatory as well as its predictive power for human decision making under risk was early queried by experimental analysts, both from the economic and psychological

disciplines (cf., e.g., ALLAIS, 1953; ELLSBERG, 1961; KAHNEMAN and TVERSKY, 1979). The deviations in behavior – compared to conventional rational-choice predictions – that were identified and replicated in numerous experiments made an important contribution to the generation of new hypotheses and the specification of theory. But even without consolidated theory, experiments may play an important role. Within policy analysis, for example, they can be used to forecast the behavioral changes that are likely to be induced by institutional innovations. Such what-if analyses (conditional forecasts) are systematic tests of how the members of a group of interest *would* decide *if* they *were* faced with a new institutional framework. In other words, understanding the black box of individual decision making is not an imperative requirement within an experimental assessment of the likely behavioral outcomes of institutional innovations such as new regulatory measures. In agricultural policy, where the mitigation of externality problems (environmental and climate protection, production of public goods, etc.) through apt regulatory mechanisms (incentives, steering taxes, allocation of novel property rights, mandatory rules) is of great importance, experimental approaches possibly open up new opportunities for a reliable forecast of farmers' behavior under changed conditions.

Since the mid-20th century, economic experiments are used to explore human decision-making behavior (cf., e.g., CHAMBERLIN, 1948; FOURAKER et al., 1962; SMITH, 1962; FOURAKER and SIEGEL, 1963). Early economic experiments dealing with the problem of public goods can be attributed to BOHM (1972). The importance of these approaches for agricultural economics results from the fact that environmental goods often have properties of public or common goods. Using field experiments and role playing games, OSTROM (1990) examines various institutions for the protection of common goods. Markets and the design of auctions (cf., e.g., ROTH, 1988) are further fields of application for experimental studies. This is due to the fact that unintended consequences and high costs for auctioneers (e.g., reduction of state revenues) may arise if people's behavior unexpectedly deviates from narrow rational-choice predictions.

The first controlled experiments were conducted with students in classroom settings.[1] The psycholo-

gists Kahneman and Tversky, for example, became famous for their classroom studies (cf., e.g., KAHNEMAN and TVERSKY, 1979; TVERSKY and KAHNEMAN, 1992). With the growing availability and computational power of modern PCs, computer-based experiments in laboratory environments (economic experimental laboratories) became increasingly popular. PCs were found to be particularly well-suited for computationally complex multi-period experiments as the considerable time delays of manual calculations could be avoided. Furthermore, partition panels for the temporary separation of the individual workstations in the lab and other technical features allowed experimenters to control the mutual influence of participants and many other environmental conditions. Even more so than in classroom experiments, a high internal validity can be obtained in well-equipped computer labs since environmental factors can be controlled with little effort. This comes at costs, however. One particular disadvantage of lab experiments is their dubious external validity caused by the fact that experimental subjects do not only face artificial decision environments but may also have to cope with unfamiliar tasks (cf., e.g. LOEWENSTEIN, 1999; LEVITT and LIST, 2007).

Instead of classroom and laboratory experiments, social psychologists adopted online experiments back in the mid-1990's (cf., e.g., REIPS, 1996; KRANTZ et al., 1997). At the same time, economic experimenters first started using online approaches to investigate various issues, such as intertemporal decisions (cf., e.g., ANDERHUB et al., 2000) or behavior, in online auctions (cf., e.g., ROTH and OCKENFELS, 2002). In recent years, online experiments have become increasingly widespread, mainly due to the fact that experimental subjects enjoy much more flexibility and are free, for example, from spatial and temporal restrictions such as having to travel to the lab at a given date (cf., e.g., ANDERHUB et al., 2001; CHARNESS et al., 2013).

Since the early 1980's, the experimental investigation of farmers' decision-making behavior under risk has been a key issue among agricultural economists (cf., e.g., BINSWANGER, 1981; REYNAUD and COUTURE, 2010; BRICK et al., 2012). Other issues have also been taken up. TRENKEL (2005), for example, discusses the advantages and disadvantages of economic experiments and points out that experiments can be used to evaluate the predictive power of rational choice predictions. He does not refer to online experiments, however. BREUSTEDT et al. (2008) experimentally investigate how various auction designs in-

---

[1] This is why the term "classroom experiment" is now used to describe an experiment that is conducted at a defined location but without using specific lab equipment (cf., e.g., KAHNEMAN et al., 1990).

fluence the willingness of farmers to participate in environmental programs. STEINHORST and BAHRS (2012) experimentally quantify the risk attitudes of both farmers and traders of agricultural goods and then analyze the consistency of their experimental choices with their risk attitudes. MUßHOFF et al. (2013) experimentally examine whether the real options approach is more suitable to explain the disinvestment decisions of farmers than the traditional net present value.

The increasing popularity of economic experiments has sparked various categorization and systematization attempts. GUALA (2005) describes important methodological aspects of experimental economics and distinguishes between lab and field experiments. The author does not concern himself, however, with the specific features and the increasing importance of online experiments. Similarly, LUSK and SHOGREN (2007) discuss the advantages and disadvantages of lab and field experiments. While pointing out that the importance of auctions has increased due to the growing importance of technology platforms such as eBay, they do not systematically contrast conventional auction experiments (for early auction experiments cf., e.g., SMITH, 1965, and SMITH, 1967) with non-online auction experiments. In contrast, the psychologist REIPS (2002) emphasizes the differences between lab and online experiments and lists some advantages and disadvantages of the latter. With a view to experimental economics, CHARNESS et al. (2007) provide a brief tabular comparison of both types of experiments. Nevertheless, the systematic classification and evaluation of online experiments remains fragmentary. Knowledge gaps persist especially regarding the question of how they can best be used for policy analysis in the agricultural sector.

With this in mind, this paper aims to systematize economic experiments and to identify the advantages and disadvantages of online experiments. Comparing conventional and online experiments, we put a specific focus on three criteria: practicability, costs, and validity. In the following section 2, we provide a comprehensive categorization of economic experiments including online experiments. In section 3, we focus on the advantages and disadvantages of Internet-based individual experiments compared to other experimental approaches. We finally discuss in section 4 the opportunities of experimental policy analysis in the agricultural sector.

2 A Categorization of Economic Experiments

The overall aim of economic experiments is to provide a better understanding of economic choices as well as reliable conditional forecasts. Experimental setups serve as systematic *tests* of how individuals (would) behave in various decision-making situations. One of the most crucial objectives of experimental studies is to identify the differences in behavior between individuals, groups, and contexts. Performance-related incentives are often considered as a necessary requirement for meaningful results in economic experiments. Performance-related incentives are to ensure that the choices that participants make in the experiment have real consequences. Such an experimental testing of people's behavior enables the analyst to combine the advantage of revealed-preference approaches with a systematic control of the environment. In other words, economic experiments allow us to generate empirical information about *real* behavior with *real* consequences under *ceteris paribus* conditions (cf., e.g., SMITH, 1982; FALK, 2001). In regular social life with its dynamic environments, in contrast, we are very rarely able to observe people's behavior under ceteris paribus conditions. The identification of behavioral determinants is thence difficult, unless we are lucky enough to stumble across a natural experiment.[2]

In an attempt to push the revelation of "true" behavior, most economists, as opposed to psychologists, incentivize experiments and grant performance-related payoffs (cf., e.g., HERTWIG and ORTMANN, 2001). Performance-related incentives are aimed to compensate participants' mental costs (*work effort*)[3], thus in-

---

[2] To avoid terminological confusion, it should be noted that surveys, even those that are explicitly designed to collect information regarding people's would-be (*hypothetical*) behavior in various situations, are not experiments. They do not reveal behavior and remain stated-preference approaches. Despite their misleading label, discrete-choice "experiments" are thus not experiments. What is labeled a "choice" is just a statement in a survey.

[3] Especially in the case of intellectually demanding experimental tasks, experimental economists usually claim that performance-related incentives must be used to compensate participants for their mental effort (*work effort hypothesis;* cf., e.g., SMITH and WALKER, 1993). Alternatively, experiments have been conducted that use social reputation as an incentive and publish the performance of participants (cf., e.g., DUERSCH et al., 2009). By contrast, psychologists often contend that extrinsic incentives and, in particular, monetary payoffs should

Figure 1.   Categorization of economic experiments
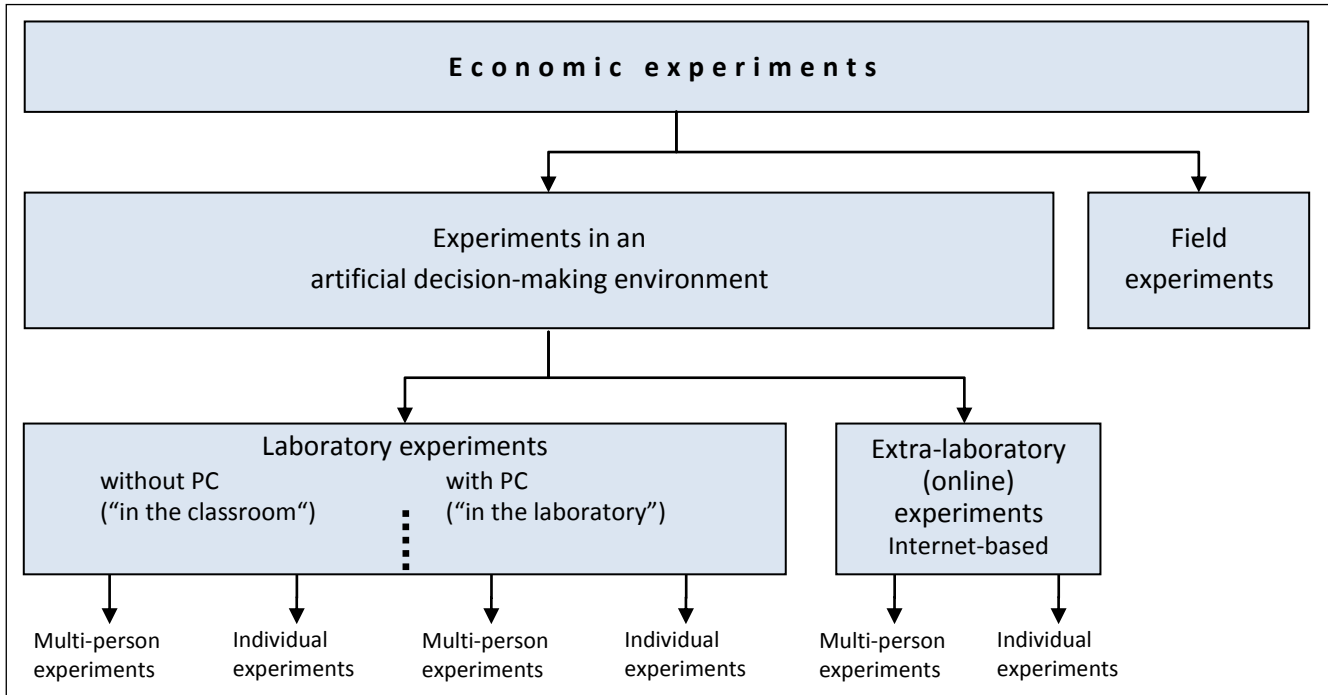
Source: own representation

ducing them to seriously consider experimental choices and to make an intellectual effort to solve the given problem. Furthermore, they aim to make participants reveal their true goals and perceptions as well as their evaluations of the potentially conflicting consequences of experimental behaviors. This includes *trade-offs* between self-interested and altruistic goals. Non-incentivized experiments may be flawed in two ways: Experimental subjects may be "lazy"; and they may exaggerate their social preferences since people have a built-in tendency to present themselves in the way they perceive to be socially desirable – at least if they can do so at no cost (*social desirability bias*; cf., e.g., MILFONT, 2009; COSTANIGRO et al., 2011; NORWOOD and LUSK, 2011).

Many different experimental designs are in line with the aforementioned general definition of economic experiments. Their precise configurations that, in turn, determine their adequateness for various research questions can differ in numerous ways. Figure 1 categorizes the different types of economic experiments.

We use four categories to classify economic experiments: First, the general type of decision-making environment in which the experiment is carried out

(experiments in an artificial environment versus field experiments); the second category is the experimental location (laboratory experiments versus extra-laboratory experiments); the third is the communication medium used (laboratory experiments without and with PC versus extra-laboratory online experiments); the fourth is the interdependencies between the participants (multi-person experiments with interdependencies versus individual experiments without interdependencies).

In general, we can distinguish between economic field experiments (controlled/randomized field trials; cf., HARRISON and LIST, 2004) and economic experiments in an artificial decision-making environment. By mimicking natural experiments, the individuals in randomized field trials, though they are actually acting in their regular living environment, are at random made subject to a change in their environmental conditions (treatment) or not. We consequently compare the treatment group with the non-treatment (or control) group to identify how treatment affects behavior. In developed industrial countries, it is hardly possible to perform true field experiments. For one thing, the costs for providing the necessary payoffs to participants in high-income countries (both for participation fees and performance-related incentives) often exceed the research budgets. Furthermore, not only moral concerns ("guinea pigs problem") but also legal constraints (equality of treatment principle) may arise that

---

not be used since they may cause unintended consequences and crowd out participants' intrinsic motivation (*crowding-out hypothesis*; cf., e.g., GNEEZY and RUSTI-CHINI, 2000).

prevent researchers from creating essentially different conditions for a random subset of the population for scientific purposes only (cf., e.g., BURTLESS, 1995; MUßHOFF and HIRSCHAUER, 2011).[4] Against this background, economic experiments are usually carried out with voluntary participants (self-selection) who make decisions in an artificially designed environment. Self-selection, however, may cause a bias (self-selection bias) that dwarfs experimental analysis (cf., e.g., ROSENTHAL and ROSNOW, 2009).

We can distinguish two groups of experiments that are conducted in an artificial decision-making environment: laboratory experiments that are performed at a defined location as opposed to extra-laboratory experiments where participants are not obliged to turn up at a certain location. Laboratory experiments can be conducted as "classroom experiments" or in an experimental laboratory equipped with PC and other technical gadgets (cf., FISCHBACHER, 1998). The great advantage of the lab is that the experimental environment can be easily controlled. Thus, experimenters can ceteris paribus isolate the behavioral influence of their variables of interest and achieve a high internal validity. However, external validity is often a problem (cf., e.g., SCHRAM, 2005) since the artificial and usually stylized experimental environment restricts the generalizability and transferability of the results to regular decision-making contexts (cf., e.g., LOEWENSTEIN, 1999). In other words, experimental economists face a trade-off between artificial environments, on the one hand, and less artificial, more realistic ("richer") environments, on the other. While the former guarantees high internal validity, the latter provides high external validity (cf., e.g., FALK, 2001). In addition, practical problems may arise to recruit enough participants for lab experiments due to the high spatial and temporal requirements that must be met by participants. Of course, this will especially be a problem in "sophisticated" approaches that attempt to recruit true representatives of the social group of interest (e.g., farmer) instead of using the conventional surrogate group of students.

Most economic studies have been conducted with "convenience" groups of students for reasons of practicability. The external validity of such studies is du-

bious if we want to learn something useful about the behavior of specific social groups other than students. With this in mind, it is hardly surprising that extra-laboratory experiments have been increasingly used in recent years. The term "extra-laboratory experiment," which has been coined by CHARNESS et al. (2013), refers in general to experiments that do not take place in the traditional lab environment. This includes studies (not depicted in Figure 1) in which individuals participate in an economic experiment, for example, in their home environment instead of in the classroom (cf., e.g., GALARZA, 2009). However, many extra-laboratory experiments in industrial countries are indeed designed as online experiments in order to reduce the burden for experimenters and facilitate participation (cf., e.g., HORTON and CHILTON, 2010).

Classroom experiments, PC-based lab experiments, and online extra-laboratory experiments can either be designed as multi-person experiments (with interdependencies between participants) or as individual experiments. Multi-person experiments with interdependencies are games in the tradition of game theory. That is, the monetary consequences (payoffs) resulting from the individual's choices do not only depend on her/his own decisions and, eventually, a stochastic environment but also on the other participants' decisions in the experiment. Exemplary game-theoretic experiments include the ultimatum game, the trust game and the public goods game (cf., e.g., LEDYARD, 1995; FEHR and GÄCHTER, 2000; MCCABE et al., 2003). In individual experiments, the experimental subject's payoffs depend exclusively on her/his choices and, eventually, a stochastic environment. Individual choices do not affect, however, the payoffs of other participants. Individual experiments can take various forms. *Simple* business simulation games in which the participants "play against the computer" and can earn money depending on their individual performance (one-person games)[5] are one example. Other examples are multiple-price lists such as incentivized Holt and Laury lotteries (cf., HOLT and LAURY, 2002) which are used to make individuals reveal their risk attitude (cf., e.g., HARRISON and RUTSTRÖM, 2008) by making them decide between several corresponding lottery pairs.

Rational choice is the axiomatic assumption of game theory that, based on formal modeling, is geared

---

[4] However, legal pilot projects that are carried out in the framework of policy or technological impact assessments can be categorized as controlled field experiments, the transferability of which is evaluated by an explicit concomitant research (cf., e.g., BJÖRKMAN and SVENSSON, 2009).

[5] It should be noted that despite their label, one-person *games* are not games in the sense of game theory since no interdependencies exist between participants.

towards the identification of strategic equilibria for exclusively self-interested and payoff-maximizing agents. While some multi-person experiments address the question of whether game-theoretic equilibria can be approximated (cf., e.g., CAMERER, 2003), others explicitly ask the question of how material payoffs in conjunction with social outcomes (social distinction, social recognition, social disrespect) and the consistency of individual choice with internalized values and identity (cf., e.g., AKERLOF and KRANTON, 2010) affect decision making (cf., e.g., CHAUDHURI, 2011). Key concerns in the latter type of analysis are the influences of positive and negative reciprocity, altruistic punishment, control aversion, and trust (cf., e.g., CHARNESS and HARUVY, 2002). Specific experimental designs enable the analyst to study the behavioral effects that are produced by social interaction, both with other individuals who actively take part in the experiments and with non-involved third parties. In this line, numerous variations of the ultimatum game (originally designed by GÜTH et al., 1982), for example, have been designed so far. They range from an anonymous implementation of the experiment without interaction between participants (cf., e.g., BOLTON and ZWICK, 1995) to a public experimental set-up where not only the experimental subjects but also non-involved third parties can observe and eventually react to the behavior of individual participants (cf., e.g., MAGEN, 2005).

3 Advantages and Disadvantages of Internet-based Individual Experiments

If a research question lends itself in principle to the field of experimental economics, analysts will have to make a difficult choice between numerous experimental designs, all of which have their advantages and disadvantages. The first question that needs to be answered is which designs enable the experimenter to answer his/her specific research question. In a further step, the designs that have been found to be generally suitable for providing answers have to be checked regarding their practicability and their ability to keep within the given research budget.

The most important methodological choice is whether to use a field experiment or an experiment in an artificial decision-making environment. Compared to field experiments, the artificial environment offers many advantages, including the researcher's discretion where and when to run the experiment, the short

period of time and relatively low costs that are required to run the experiment, the improved control of the environment, and, last not least, an easy replication resulting from the aforementioned advantages. It must be noted, however, that the generation of experimental data in artificial decision-making environments also involves serious drawbacks. We always need to critically scrutinize, for example, if, or to what extent, the causalities that have been identified in the experiment can be generalized to the considerably more complex regular life environment of the social group of interest.

3.1 Practicability, Costs, and Time Required for Conducting Experiments

Online experiments can be easily carried out in developed industrial countries due to the wide distribution of modern communication technologies. Compared to lab experiments, the advantages that are inherent to the artificial experimental environment in general (flexibility, speed, costs, and easy replication) are even more pronounced for online experiments. A different picture emerges if one looks at the controllability of the experimental environment and its implications for internal validity (see section 3.2).

Online experiments have considerable advantages over lab experiments due to the spatial and temporal flexibility they offer to the participants and the speed in which experiments can be conducted. In online experiments, no travel costs arise since experimental subjects do not have to travel to a lab.[6] Furthermore, the opportunity costs of participants are greatly reduced for two reasons: First, participants enjoy a great temporal flexibility as to when exactly to take part in the experiment. Second, the total time required for participation in the experiment is lower. Both features facilitate experiments with larger numbers of participants. In addition, it increases the researcher's discretion to include experimental subjects from various regions, including high distance locations. Not only the burden for participants in online experiments but also the direct costs for experimenters are lower since no lab capacities have to be provided. This facilitates a quasi-unlimited simultaneous as well as sequential participation of large numbers of participants. Lab experiments, in contrast, always face capacity limita-

---

[6] For mobile laboratories (cf., e.g., SCHADE and BURMEISTER-LAMP, 2009), the participants' travel expenses are low or non-existent. In contrast, the experimenter's travel costs are on the rise.

tions, and several experimental runs may be needed to obtain sufficiently large numbers of participants.[7]

In individual experiments, the researcher can exploit the organizational advantages of the Internet without limitation. In multi-person experiments that are carried out in the tradition of game theory (i.e., with interdependencies between participants), the participants' autonomy, which is an inherent feature of the online approach, may cause problems. On the one hand, "technical" problems may arise if experimental subjects do not take part at the same time or if they do not complete the entire experiment. On the other, undisclosed bias and internal validity problems may be caused if they engage in mutual communication and make agreements that are not controlled by the experimenter. Problems of the latter kind are especially critical for experiments that are explicitly aimed to identify how social interaction affects human behavior.

3.2 Internal and External Validity

With the control over the decision-making environment being low in online experiments, the researcher does not exactly know which sources of information and resources the participants use and how they communicate with others during the experiment. It thus remains uncertain, for example, how much time the participants actually use for the experimental tasks. With regard to validity, "going online" in experiments has a double-edged effect. Internal validity is threatened, but the more realistic setting of the decision environment may increase external validity (i.e., the generalizability of findings towards regular life contexts). Several features of online experiments contribute to more "realism". First, participants can make their experimental choices in the familiar setting of their own home office. Second, as in regular life, they have the discretion to adjust their effort in the experiment according to the marginal benefits and costs they subjectively expect. Third, a non-intended influence of the researcher as a person is precluded in experiments, whereas such interferences are a threat in the lab where the experimenter is often personally present.

Given a fixed research budget, replacing lab experiments by online experiments provides several options to increase external validity. This is due to the overall lower costs of the online approach and the resulting free financial resources that can be used for other purposes. Besides the savings of travel and lab costs, funds are also freed in the online approach because the participants have less opportunity costs that must be compensated. In other words, the "net value" of a given participation fee is ceteris paribus higher in an online experiment than in a lab experiment.

Figure 2 describes the range of opportunities that are provided to online researchers through the lab and travel cost savings and the reduction of participants' opportunity costs. Besides simply using the cost savings to reduce the required research budgets (cf., the left box shown in grey), online experimenters who dispose of a given research budget have three relevant options (cf., the three black boxes shown on the right) to use freed funds to increase external validity:

1. They can use the freed funds (including those resulting from a feasible reduction of the participation fees) to run the experiment with a higher number of participants.

2. They can use the freed funds (including those resulting from a feasible reduction of the participation fees) to increase the performance-related incentives. This facilitates the reduction of an eventual gap between (low) experimental incentives and the actual consequences of people's actions in their regular decision environment.

3. They can use the lab and travel savings to increase the participation fee and recruit representatives of the social group of interest[8] as experimental subjects instead of students. Increased participation fees, in conjunction with the general reduction of participants' opportunity costs in online experiments, enables experimenters to better cover the opportunity costs of participants of the social group of interest (e.g., farmers) that are regularly higher than those of students who are often used in experiments as convenient surrogates.

Researchers can also employ various mixtures of the aforementioned measures and use, for example, one part of the freed funds to increase performance-related incentives and another part to recruit real decision makers from the social group of interest as experimental subjects.

---

[7] In online experiments, practical difficulties arise with anonymity and privacy issues. If experimenters collect anonymous data only in order to push the revelation of "true" behavior, they will have a difficult job to make payments to participants for their participation and, even more so, for their performance.

[8] Younger participants are usually more familiar with modern communication technology than elderly participants. The online experimenter must therefore consider and eventually reduce the risk of a bias in order to ensure the representativeness of participants.

Figure 2.   How cost savings of online experiments can be used to increase external validity
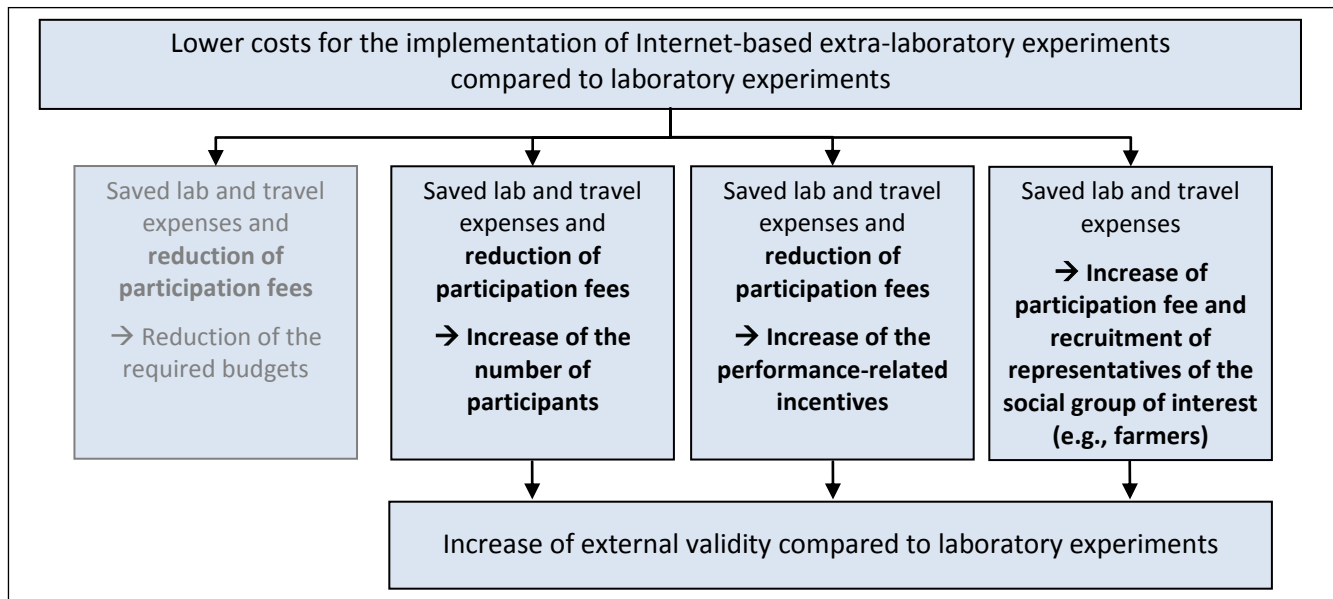
Source: own representation

An increase of the performance-related incentives helps to compensate the mental effort of participants and make them seriously attack the experimental tasks. Furthermore, incentives that are paid contingent on performance reduce the ever imminent *social desirability bias* if trade-offs between performance and pro-social goals are designed into an experiment (antagonistic incentives). In other words, if relevant payoffs are at stake, the participants cannot behave as pro-social "do-gooders" (*Gutmenschen*) without costs. However, while reducing the social desirability bias, using performance-related incentives generates the risk of another problem in individual experiments, especially if they are aimed at identifying pro-social preferences. Performance-related incentives may induce participants, even if they have pro-social goals, to adopt purely self-interested behaviors and exclusively strive for monetary payoffs in the experiment. That is, they may realize that experimental payoffs are real while the social consequences are only virtual in that no social costs for themselves or others are caused in reality. This problem which hampers the revelation of the actors' true preferences can be labeled *virtual game bias*. In addition, high performance-related incentives may increase the share of *virtual game bias* participants who, from the very start, are pre-determined to exclusively aim to earn money in the experiment. Therefore, a *self-selection bias* possibly intensifies the virtual game bias.

In individual experiments that are used to investigate the trade-offs between conflicting self-interested goals, monetary incentives increase the chance that individuals reveal their true preferences without causing the risk of a virtual game bias. One example are Holt and Laury experiments which investigate how individuals assess the trade-offs between income and risk (risk premium). Virtual game bias is not a problem either in individual experiments that are designed to analyze bounded rationality (e.g., framing effects). However, when individual experiments investigate the trade-offs between self-interested goals and pro-social goals (ethical premium), the virtual game bias cannot be completely avoided. To obtain meaningful empirical evidence in individual experiments, the participants must be randomly assigned to experimental scenarios that are identical regarding the monetary incentives (real monetary consequences attached to individual action) but differ with respect to the communicated social consequences. *Within groups*, experimental subjects will be heterogeneous regarding the revelation of their true preferences. Thus, we cannot observe the level of their ethical premium (i.e., their willingness to renounce profits in exchange for the achievement of pro-social goals) since we do not know to which extent the experimentally observed behavior is distorted by the virtual game bias. The differences *between groups* are meaningful, however. In a cautious interpretation they can be understood as lower bounds of the ethical premium.

4 Future Prospects

Economic experiments are run for many reasons, including the generation of hypotheses and theory testing. The field of experimental applications is not restricted to positive analysis, but also covers normative analysis for social actors who attempt to steer the behavior of others. In policy analysis, economic experiments could be used to assess the behavioral changes that are likely to be induced by institutional innovations. Such tests of behavior and their prudent interpretation as conditional forecasts (what-if analyses) do not require the experimenter to fully understand the black box of individual decision making.

The importance of a reliable policy impact assessment particularly results from recent changes in European agricultural policy. Agricultural policy increasingly attempts to steer the behavior of economic actors through changing their institutional environment, instead of granting transfer payments which was the predominant approach in the past (cf., e.g., VERCAMMEN, 2011). On the one hand, new policies are aimed at reducing negative externalities such as the social costs caused by nitrogen entry into the groundwater, emission of climate-damaging gases, pesticide residues in food, etc. On the other hand, they try to solve positive externality problems, such as the underproduction of public goods, which are deemed socially desirable but not remunerated in the conventional market environment (e.g., cultural landscapes and biodiversity, animal welfare, biological carbon dioxide stores, etc.). Any cost-benefit analysis that is based on socially desired goals within a framework of policy analysis first of all needs to identify measures that are cost-efficient with regard to the intended change of actors' behavior. Naturally, in this endeavor, the regulator has to identify the most likely would-be behavior of the actors concerned.

In general, the steering effect of institutional innovations can be analyzed ex-post or ex-ante. Ex-post approaches evaluate measures that have been implemented in the actor's real environment. A major drawback of ex-post approaches is that the budgets have been spent already. In addition, there is a lack of internal validity since empirically observed socio-economic phenomena cannot easily be attributed to a certain policy measure due to the limited control of the relevant environment. Thus, ex-post approaches often are restricted to the identification of statistical relations (e.g., econometric analyses) or comparative analyses (e.g., systematic case studies).

Ex-ante approaches have the advantage of providing decision support before political choices are made. However, often narrow rational choice models, which assume a completely rational and exclusively profit-maximizing agent (homo oeconomicus), have been used in ex-ante policy analysis. Real individuals, however, pursue multiple goals. They do not exclusively strive for profit and risk reduction but also for non-monetary goals including social recognition and consistency with internalized values and identity. Furthermore, people in the real world are bounded-rational decision makers. Being limited in their cognitive abilities as well as in their information about the relevant environment, they often rely on simple decision-making rules and heuristics (cf., SIMON, 1957). Using rational choice approaches thus causes the risk of misjudging both the type and speed of adaptive behavior to changes in their environment. Formal utility models, which extend the narrow rational choice perspective by multiple goals and bounded rationality, are a first attempt to avoid such misjudgments (cf., e.g., FEHR and SCHMIDT, 1999). However, economic actors are heterogeneous in their goals, evaluations, and the extent of their bounded rationality. Conditional forecasts for the decision-making behavior of individual actors with the help of formal models are usually not viable due to lack of subject-related data or the prohibitively high costs of collecting such data. In most cases, formal models are therefore not suited to shed light on what is often misleadingly referred to as "behavioral anomaly", i.e., the gap between forecasts based on (narrow) rational choice and actual behavior. Economic experiments, in contrast, facilitate a systematic and context-related testing of human behavior with real consequences under controlled environmental conditions.

In recent years, business simulation games have been increasingly suggested as an instrument to experimentally provide conditional forecasts for policy impact analysis. In business simulation games, the participants run, for example, a farm and make decisions about their production program. By controlling the rules of the game and a systematic ceteris paribus variation of factors (e.g., policy measures) that are assumed to affect behavior, causal relationships can be identified. This, in turn contributes to the specification of theory concerned with human decision making. The decision-making context for business simulation games can be designed close to reality, independent of whether they are conducted in the lab or online. This improves their external validity compared to more

stylized experimental contexts. The general "external validity gap" of experiments can be further reduced if experimental subjects are recruited from the social group of interest (e.g., with farmers) instead of using surrogates such as students. This will be important if, in agricultural policy analysis, we want to learn something about the behavior of farmers (and not of students) given a specific institutional innovation. Within the framework of a systematic methodology triangulation, the experimental testing of the *real* behavior of *real* decision makers who face relevant *real* payoffs in a controlled and artificial but realistic decision-making environment may complement more traditional rational choice approaches. Systematic triangulation, in turn, has a good chance to increase the quality of conditional behavioral forecasting.

While first approaches to use experiments for agricultural policy analysis are available (cf., e.g., BREUSTEDT et al., 2008; MUßHOFF and HIRSCHAUER, 2011), their general suitability for specific questions of agricultural policy design and, eventually, the need for methodological specifications need to be further studied. If experimental approaches can be confirmed to be a suitable tool to provide reasonably reliable forecasts of actors' behaviors, a vast variety of subject areas lends itself to an analysis. Given the pollution of ground and surface waters resulting from the nitrogen surplus of nearly 100 kg/ha (in 2011 according to UBA, 2014), a prominent field of application would be the question of how various policy measures affect farmers' fertilizing behavior. The question of how policy choices affect behavior is especially relevant if policy makers enlarge their toolbox and think not only beyond traditional command and control approaches (e.g., mandatory fertilizing limits) but also beyond traditional market approaches such as a change in relative prices (e.g., by means of a steering tax).

References

AKERLOF, G. and R. KRANTON (2010): Identity economics. Princeton University Press, Princeton.

ALLAIS, M.P. (1953): Le Comportement de l'Homme Rationnel devant le risque: Critique des postulates et axiomes de l'Ecole Americaine. In: Econometrica 21 (4): 503-546.

ANDERHUB, V., W. GÜTH, W. MÜLLER and M. STROBEL (2000): An Experimental Analysis of Intertemporal Allocation Behavior. In: Experimental Economics 3 (2): 137-152.

ANDERHUB, V., R. MÜLLER and C. SCHMIDT (2001): Design and Evaluation of an Economic Experiment via the Internet. In: Journal of Economic Behavior & Organization 46 (2): 227-247.

BINSWANGER, H.P. (1981): Attitudes toward risk: theoretical implications of an experiment in rural India. In: The Economic Journal 91 (364): 867 (-890)

BJÖRKMAN, M. and J. SVENSSON (2009): Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda. In: Quarterly Journal of Economics 124 (2): 735-769.

BOHM, P. (1972): Estimating Demand for Public Goods: An Experiment. In: European Economic Review 3 (2): 111-130.

BOLTON, G.E. and R. ZWICK (1995): Anonymity versus Punishment in Ultimatum Bargaining. In: Games and Economic Behavior 10 (1): 95-121.

BREUSTEDT, G., U. LATACZ-LOHMANN and S. SCHILIZZI (2008): Ein ökonomisches Auktionsexperiment zur Auswahl der Teilnehmer an Umweltschutzprogrammen. Agrar- und Ernährungswirtschaft im Umbruch. Schriften der GEWISOLA 43, Münster: 41-49.

BRICK, K., M. VISSER and J. BURNS (2012): Risk Aversion: Experimental Evidence from South African Fishing Communities. In: American Journal of Agricultural Economics 94 (1): 133-152.

BURTLESS, G. (1995): The Case for Randomized Field Trials in Economic and Policy Research. In: The Journal of Economic Perspectives 9 (2): 63-84.

CAMERER, C.F. (2003): Behavioral game theory. Experiments in strategic interaction. Princeton University Press, Princeton.

CHAMBERLIN, E.H. (1948): An Experimental Imperfect Market. In: Journal of Political Economy 56 (2): 95-108.

CHARNESS, G., U. GNEEZY and M.A. KUHN (2013): Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics. In: Journal of Economic Behavior & Organisation 91: 93-100.

CHARNESS, G. and E. HARUVY (2002): Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach. In: Games and Economic Behavior 40 (2): 203-231.

CHARNESS, G., E. HARUVY and D. SONSINO (2007): Social distance and reciprocity: An Internet experiment. In: Journal of Economic Behavior & Organization 63 (1): 88-103.

CHAUDHURI, A. (2011): Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. In: Experimental Economics 14 (1): 47-83.

COSTANIGRO, M., D.T. MCFADDEN, S. KROLL, and G. NURSE (2011): An In-Store Valuation of Local and Organic Apples: The Role of Social Desirability. In: Agribusiness 27 (4): 465-477.

DUERSCH, P., J. OECHSSLER and B.C. SCHIPPER (2009): Incentives for subjects in internet experiments. In: Economics letters 105 (1): 120-122.

ELLSBERG, D. (1961): Risk, Ambiguity and the Savage Axioms. In: Quarterly Journal of Economics 75 (4): 643-669.

FALK, A. (2001): Homo oeconomicus versus Homo reciprocans: Ansätze für ein neues wirtschaftspolitisches Leitbild? Working Paper 79. Institut für Empirische Wirtschaftsforschung, Universität Zürich.

FEHR, E. and S. GÄCHTER (2000): Cooperation and Punishment in Public Goods Experiments. In: The American Economic Review 90 (4): 980-994.

FEHR, E. and M. SCHMIDT (1999): A theory of fairness, competetition, and cooperation. In: The Quarterly Journal of Economics 114 (3): 817-868.

FISCHBACHER, U. (1998): z-Tree: Zurich Toolbox for Readymade Economic Experiments. Instructions for Experimenters. Mimeo. University of Zurich.

FOURAKER, L.E. and S. SIEGEL (1963): Bargaining Behavior. McGraw Hill, New York.

FOURAKER, L.E., S. SIEGEL and D. HARNETT (1962): An Experimental Disposition of Alternative Bilateral Monopoly Models under Conditions of Price Leadership. In: Operations Research 10 (1): 41-50.

GALARZA, F.B. (2009): Choices under risk in rural Peru. MPRA Paper 17708. University Library Munich.

GNEEZY, U. and A. RUSTICHINI (2000): Pay Enough or Don't Pay at All. In: Quarterly Journal of Economics 155 (3): 791-810.

GUALA, F. (2005): The Methodology of Experimental Economics. Cambridge University Press, Cambridge.

GÜTH, W., R. SCHMITTBERGER and B. SCHWARZE (1982): An experimental analysis of ultimatum bargaining. In: Journal of Economic Behavior and Organisation 3 (4): 367-388.

HARRISON, G.W. and J.A. LIST (2004): Field Experiments. In: Journal of Economic Literature 42 (4): 1009-1055.

HARRISON, G.W. and E.E. RUTSTRÖM (2008): Risk aversion in the laboratory. In: Research in Experimental Economics 12: 41-196.

HERTWIG, R. and A. ORTMANN (2001): Experimental Practices in Economics: A Methodological Challenge for Psychologists? In: Behavioral and Brain Sciences 24 (3): 383-451.

HOLT, C.A. and S.K. LAURY (2002): Risk Averson and Incentive Effects. In: The American Economic Review 92 (5): 1644-1655.

HORTON, J.J. and L.B. CHILTON (2010): The labor economics of paid crowdsourcing. Proceedings of the 11th ACM Conference on Electronic Commerce, New York.

KAHNEMAN, D., J.L. KNETSCH and R.H. THALER (1990): Experimental Tests of the Endowment Effect and the Coase Theorem. In: The Journal of Political Economy 98 (6): 1325-1348.

KAHNEMAN, D. and A. TVERSKY (1979): Prospect Theory: An Analysis of Decision under Risk. In: Econometrica 47 (2): 263-292.

KRANTZ, J.H., J. BALLARD and J. SCHER (1997): Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. In: Behavior Research Methods, Instruments, & Computers 29 (2): 264-269.

LEDYARD, J.O. (1995): Public Goods: A Survey of Experimental Research. In: Kagel, J.H. and A.E. Roth (eds.): Handbook of Experimental Economics. Princeton University Press, Princeton: 111-194.

LEVITT, S.D. and J.A. LIST (2007): What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? In: Journal of Economic Perspectives 21 (2): 153-174.

LOEWENSTEIN, G. (1999): Experimental economics from the vantage-point of behavioural economics. In: The Economic Journal 109 (453): F25-F34.

LUSK, J.L. and J.F. SHOGREN (2007): Experimental Auctions: Methods and Applications in Economic and Marketing Research. Cambridge University Press, Cambridge.

MAGEN, S. (2005): Fairness, Eigennutz und die Rolle des Rechts: Eine Analyse auf Grundlage der Verhaltensökonomik. Preprint des Max-Planck-Instituts zur Erforschung von Gemeinschaftsgütern Nr. 2005/22, Bonn.

MCCABE, K.A., M.L. RIGDON and V.L. SMITH (2003): Positive reciprocity and intentions in trust games. In: Journal of Economic Behavior & Organization 52 (2): 267-275.

MILFONT, T.L. (2009): The effects of social desirability on self-reported environmental attitudes and ecological behaviour. In: Environmentalist 29 (3): 263-269.

MUßHOFF, O. and N. HIRSCHAUER (2011): Bereitstellung ökosystemarer Dienstleistungen – Eine experimentelle Folgenabschätzung politischer Steuerungsinstrumente der Stickstoffreduzierung. In: Zeitschrift für Umweltpolitik & Umweltrecht (ZfU) 4/2011: 437-468.

MUßHOFF, O., M. ODENING, C. SCHADE, S.C. MAART and S. SANDRI (2013): Inertia in Disinvestment Decisions: Experimental Evidence. In: European Review of Agricultural Economics 40 (3): 463-485.

NORWOOD, F.B. and J.L. LUSK (2011): Social Desirability Bias in Real, Hypothetical and Inferred Valuation Experiments. In: American Journal of Agricultural Economics 93 (2): 528-534.

OSTROM, E. (1990): Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press, New York.

REIPS, U.D. (1996): Experimentieren im World Wide Web. In: Experimentelle Psychologie: Tagung experimentell arbeitender Psychologen 38: 256-257.

– (2002): Standards for Internet-based experimenting. In: Experimental Psychology 49 (4): 243-256.

REYNAUD, A. and S. COUTURE (2010): Stability of risk preference measure: results from a field experiment on french farmers. In: Theory and Decision 73 (2): 203-221.

ROSENTHAL, R. and R.L. ROSNOW (2009): The Volunteer Subject. In: R. Rosenthal and R.L. Rosnow (eds.): Artifacts in Behavioral Research. Robert Rosenthal and Ralph L. Rosnow's Classic Books. Oxford University Press, Oxford: 48-92.

ROTH, A. (1988): Laboratory experimentation in economics: Six points of view. Cambridge University Press, Cambridge.

ROTH, A. and A. OCKENFELS (2000): Last Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon Auctions on the Internet. In: American Economic Review 92 (4): 1093-1103.

SCHADE, C. and K. BURMEISTER-LAMP (2009): Experiments on entrepreneurial decision making: A different lens through which to look at Trends in Entrepreneurship. In: Foundations and Trends in Entrepreneurship 5 (2): 81-134.

SCHRAM, A. (2005): Artificiality: The tension between internal and external validity in economic experiments. In: Journal of Economic Methodology 12 (2): 225-237.

SIMON, H.A. (1957): Models of Man: Social and Rational. Wiley, New York.

SMITH, V.L. (1962): An Experimental Study of Competitive Market Behavior. In: Journal of Political Economy 70 (2): 111-137.

– (1965): Experimental Auction Markets and the Walrasian Hypothesis. In: Journal of Political Economy 73 (4): 387-393.

– (1967): Experimental Studies of Discrimination Versus Competition in Sealed-Bid Auction Markets. In: The Journal of Business 40 (1): 56-84.

– (1982): Microeconomic Systems as an Experimental Science. In: The American Economic Review 72 (5): 923-955.

SMITH, V.L. and J.M. WALKER (1993): Monetary Rewards and Decision Cost in Experimental Economics. In: Economic Inquiry 31 (2): 245-261.

STEINHORST, M.P. and E. BAHRS (2012): Die Analyse der Rationalität im Verhalten von Stakeholdern des Agribusiness anhand eines Experiments. Unternehmerische Landwirtschaft zwischen Marktanforderungen und gesellschaftlichen Erwartungen. Schriften der GEWISOLA 47, Münster: 165-176.

TRENKEL, H. (2005): Möglichkeiten experimenteller Methoden in der Agrarökonomie. Schriften der GEWISOLA 40, Münster: 447-453.

TVERSKY, A. and D. KAHNEMAN (1992): Advances in Prospect Theory: Cumulative Representation of Uncertainty. In: Journal of Risk and Uncertainty 5 (4): 297-323.

UBA (Umweltbundesamt) (2014): Nährstoffeinträge aus der Landwirtschaft und Stickstoffüberschuss. In: http://www.umweltbundesamt.de/daten/land-forstwirt schaft/landwirtschaft/naehrstoffeintraege-aus-der-land wirtschaft (last accessed on 25th February 2014).

VERCAMMEN, J. (2011): Agri-Environmental Regulations, Policies, and Programs. In: Canadian Journal of Agricultural Economics 59 (1): 1-18.

VON NEUMANN, J. and O. MORGENSTERN (1944): Theory of Games and Economic Behavior. Princeton University Press, Princeton.

Acknowledgement

Contact author:
DR. SVEN GRÜNER
Martin-Luther-Universität Halle-Wittenberg
Institut für Agrar- und Ernährungswissenschaften
Karl-Freiherr-von-Fritsch-Str. 4, 06120 Halle (Saale)
e-mail: sven.gruener@landw.uni-halle.de