



Behavioral and Experimental Agri-Environmental Research: Methodological Challenges, Literature Gaps, and Recommendations

Leah H. Palm-Forster¹ · Paul J. Ferraro² · Nicholas Janusch³ · Christian A. Vossler⁴ · Kent D. Messer¹

Accepted: 19 March 2019 / Published online: 3 April 2019
© Springer Nature B.V. 2019

Abstract

Insights from behavioral and experimental economics research can inform the design of evidence-based, cost-effective agri-environmental programs that mitigate environmental damages and promote the supply of environmental benefits from agricultural landscapes. To enhance future research on agri-environmental program design and to increase the speed at which credible scientific knowledge is accumulated, we highlight methodological challenges, identify important gaps in the existing literature, and make key recommendations for both researchers and those evaluating research. We first report on four key methodological challenges—underpowered designs, multiple hypothesis testing, interpretation issues, and choosing appropriate econometric methods—and suggest strategies to overcome these challenges. Specifically, we emphasize the need for more detailed planning during the experimental design stage, including power analyses and publishing a pre-analysis plan. Greater use of replication studies and meta-analyses will also help address these challenges and strengthen the quality of the evidence base. In the second part of this paper, we discuss how insights from behavioral and experimental economics can be applied to improve the design of agri-environmental programs. We summarize key insights using the MINDSPACE framework, which categorizes nine behavioral effects that influence decision-making (messenger, incentives, norms, defaults, salience, priming, affect, commitment, and ego), and we highlight recent research that tests these effects in agri-environmental contexts. We also propose a framework for prioritizing policy-relevant research in this domain.

Keywords Behavioral insights · Conservation · Effect size · Environmental economics · Experimental design · Power analysis · Subject pools

JEL Classification C9 · D9 · Q58 · Q52

✉ Leah H. Palm-Forster
leahhp@udel.edu

Extended author information available on the last page of the article

1 Introduction

Although many developed and developing countries have made considerable strides in improving environmental quality and conserving natural resources, the objectives of government environmental protection agencies are not universally met. For example, in the United States, over 53% of stream and river miles and 70% of lake, reservoir, and pond acres are listed as “impaired”, meaning that current water quality levels fall below designated use standards (U.S. Environmental Protection Agency 2018). As approximately 37% of the world’s land area is agricultural land (FAO 2016), one mechanism for improving environmental quality, and the production of ecosystem services (e.g., pollination services, habitat provision) more broadly, is funding agri-environmental programs.

Agri-environmental programs promote the use of agricultural best management practices (BMPs; also referred to as “beneficial management practices”), or otherwise encourage conservation efforts and motivate environmental stewardship. The success of these programs depends on how effectively limited resources can be used to motivate agricultural producers to change their behavior in ways that support the provision of environmental benefits. To more cost-effectively motivate these behavioral changes, program managers have expressed a growing interest in applying insights from behavioral and experimental economics research (Higgins et al. 2017). To facilitate these applications, this paper seeks to summarize the state of behavioral and experimental economics research in the agri-environmental context. In doing so, we aim to enhance future research on agri-environmental programs and to increase the speed at which credible scientific knowledge is accumulated in this area.

To meet our objectives, we draw upon insights generated through the Conference on Behavioral and Experimental Agri-environmental Research: Methodological Advancements and Applications to Policy (CBEAR-MAAP), which was held in October 2017. This conference was organized by the Center for Behavioral and Experimental Agri-environmental Research (CBEAR) and funded by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture. The conference showcased experimental and behavioral economics research addressing agri-environmental management and policy challenges. This Special Issue of *Environmental & Resource Economics* contains, in revised form, a selection of papers that were presented at the conference.

Much of the discussion during the conference focused on methodological challenges when conducting agri-environmental research using experimental methods. Conference participants also discussed the state of the literature and research that could improve the design of programs and policies. In this article, we: (1) examine four key methodological challenges that were discussed during the conference and suggest strategies to overcome them; (2) review the behavioral and experimental economics literature using the Ag-E MINDSPACE framework (based on the framework developed by Dolan et al. 2012), which categorizes behavioral insights that apply to agri-environmental research; and (3) provide recommendations on how to improve future research that will inform policy and program design.

2 Methodological Issues Related to the Application of Experimental Designs in the Agri-Environmental Context

Experimental economics methods have been an essential tool for environmental economists for decades (Cherry et al. 2008), and laboratory and field experiments are increasingly used to inform policies and programs that seek to address agri-environmental

challenges (Higgins et al. 2017). As with broader environmental issues, agri-environmental challenges often involve collective action problems in which individuals behaving in their own self-interest limit social efficiency (Sturm and Weimann 2006; Wallander et al. 2017). Examples include mitigating negative externalities from agricultural production, the provision of environmental services (public goods), and common-pool resource management. Economics experiments are valuable research tools for studying such problems because they permit direct measurement of individual preferences (e.g., risk, time, pro-social, environmental) and characteristics (e.g., trust, reciprocity) that are not typically observable, but greatly impact agri-environmental decision making. Furthermore, by controlling the decision environment and exogenously varying the key interventions (treatments), researchers seeking to draw causal inferences confront fewer non-causal explanations for observed patterns in the data relative to other empirical approaches.

In the past decade, there has been a growing interest in expanding and improving experimental methodologies in environmental research, including papers and edited volumes by prominent researchers (Harrison and List 2004; Cherry et al. 2008; Ehmke and Shogren 2009; Messer et al. 2014; List and Price 2016; Cason and Wu 2018). We contribute to this literature by highlighting methodological challenges discussed during the CBEAR-MAAP conference, which can be separated into issues associated with: (1) underpowered designs; (2) multiple hypothesis testing; (3) interpretation of results; and (4) econometric methods. We also propose strategies for addressing these challenges.

2.1 Methodological Challenge #1: Underpowered Designs

Statistical power refers to the probability that a study design will reject a false null hypothesis (i.e., not make a Type II error).¹ In practice, a common rule of thumb is to aim for 80% power or better (i.e., a 20% chance of making a Type II error). Economists often fail to conduct *ex ante* power analyses. Ioannidis et al. (2017) survey the economics literature and report an alarming result: the median statistical power is 18%. Most of the economic studies included in their analysis relied on observational data, but similar concerns have been raised for experimental economics studies (Zhang and Ortmann 2013), and experiments in the behavioral sciences (Smaldino and McElreath 2016).

Not only does the implementation of underpowered designs lead to false negatives, it increases the chances of committing what are known as Type M errors and Type S errors (Button et al. 2013; Gelman and Carlin 2014). A Type M error is an error of magnitude: a claim that the treatment effect is small in magnitude when it is, in fact, large, or vice versa. A Type S error is an error of sign: a claim that an effect is negative when it is, in fact, positive, or vice versa. With an underpowered design, the sampling distribution of sample means is wide (high variance) and thus more extreme estimates are likely. Thus, if an estimate is statistically significant in an underpowered design, it is very likely to be larger in magnitude than the true effect size and also has a good chance of being the wrong sign.

If all studies were published and replications were common, Type M and Type S errors would not be a serious problem. However, there may be bias in the peer-review process towards publishing statistically significant results (Doucouliagos and Stanley 2013). Second, a similar bias against publishing replication studies means that experiments,

¹ Ellis (2010) provides a helpful review of concepts including statistical power, effect sizes, and meta-analysis.

particularly expensive field experiments, are rarely replicated. These two biases limit our ability to identify exaggerated and spurious estimates reported in the literature. In the same analysis that estimated the median power in economics studies as 18%, the authors also estimated that 80% of reported effects are exaggerated, typically by a factor of two or more.

Many experimental designs investigating agri-environmental issues, including prior studies conducted by the authors, are likely to be underpowered. Underpowered designs are particularly pervasive in investigations of behavioral nudges in field experiments, where the true effect sizes are relatively small and require very large sample sizes to detect. When the true effect is positive, the most likely inferences from such experiments are that the intervention is unsuccessful or has a negative effect, or a positive effect that is exaggerated. Such inferences can thwart future refinements of interventions, prevent more widespread adoption of successful interventions, and encourage widespread adoption of interventions with limited (or even negative) impacts. In other words, underpowered designs, particularly in the presence of publication biases and the absence of replications, can encourage the misallocation of scarce agri-environmental program resources.

To help identify underpowered designs in advance, researchers should conduct power analyses prior to implementing an experiment (Ellis 2010; List et al. 2011).² Power analyses also help reduce the likelihood of other misguided and problematic practices, such as endogenously choosing the sample size based on the results obtained after collecting data. Without a sample size established in advance, and with the challenges associated with publishing null results, researchers may be tempted to continue collecting data as long as the results are statistically insignificant, and then stop after obtaining a statistically significant result. Using statistical significance to determine whether to stop or continue a study has long been known to introduce bias into empirical study designs (Armitage et al. 1969). This decision rule leads to a high probability of committing a Type I error (i.e. rejecting a true null hypothesis), which necessarily exceeds the claimed significance levels reported in the studies (Simmons et al. 2011). Sample sizes should be determined in advance. Given that researchers often have difficulty publishing null results, an additional benefit from having a sufficiently powered design is that it mitigates reviewers' concerns about the lack of information contained in a null result, as researchers can point to the power analysis that is part of the publicly-available pre-analysis plan and the magnitude of effect that the study was designed to detect.³

Power calculations depend on many factors, including sample size, the assumed Type I error rate, the variance of the outcome variable, the effect size, and the testing procedure employed (e.g. econometric model or choice of simple parametric or nonparametric test). Other facets of the design, such as whether there are repeated measurements from the same observational units, and whether treatments are implemented within or between subjects, also need to be considered. Bellemare et al. (2016) provide an overview of available software programs and packages available for conducting power analyses. Moreover,

² Lakens et al. (2018) emphasize that researchers should also justify their alpha level (i.e. the statistical significance level) along with other decisions when designing a study.

³ See Brown et al. (2019) for a discussion about statistical methods that can be used to interpret null findings, which can be meaningful and policy-relevant when derived from a well-designed study.

they introduce a simulation-based module for Stata that is specifically design for economics experiments.^{4,5}

A primary challenge is to specify the expected outcome distributions associated with the treatment and control conditions. Of course, if we knew these distributions precisely, we would not have to conduct the experiment! One approach is to use theory or expert opinion to guide our estimates. Another approach is to use data collected through a pilot study, using a sample from the population of interest. These data can often be helpful in providing an estimate of the outcome distribution. Moreover, pilots can often be useful for other reasons, such as refining experiment parameters, procedures, experiment software, and information materials.

As an alternative or complementary approach, researchers can rely on meta-analyses that “generate a pooled estimate of an effect size that accounts for the sampling and measurement error attached to individual estimates” (Ellis 2010, p. 61). For example, meta-analyses can be used to estimate *standardized* effect sizes, which give researchers a sense of the relative impact of an intervention. They are particularly useful when making comparisons across studies that may have different outcome measures or employ measures based on arbitrary or differing scales. One standardized effect size measure is Cohen’s *d* (Cohen 1988), defined as the difference between two means (i.e. the treatment effect) divided by the pooled standard deviation of the two samples.^{6,7} Thalheimer and Cook (2002) provide a clear guide of how to calculate these standardized effect sizes, including cases where one instead uses information provided through *t* and *F* statistics rather than standard deviations. When usable estimates from meta-analyses are not available, researchers can pool individual estimates from prior publications.

Researchers drawing on meta-analyses or prior publications should, however, remember that prior studies may be subject to publication biases and be underpowered. Thus, their reported effect sizes may be much higher than the true effect sizes. Many field studies related to agri-environmental issues report an effect size of just 0.10 standard deviation or less.⁸ Some experiments need to be designed to detect even smaller effects. For example, low-cost program interventions, such as subtle wording changes in agri-environmental program materials, may have small expected effect sizes, but are worth testing because they can be cost-effective even with a small effect size. Therefore, we recommend that researchers err towards the side of caution, and design experiments to achieve minimum detectable effect sizes that are conservative (small) relative to values recorded in related research.

There are at least three additional obstacles to using power analysis to inform study designs. First, researchers must be willing to accept the results of the power analysis and

⁴ This module accommodates within and between-subject experimental designs that may involve complications such as continuous treatment variables, order effects, and repetition. Further, power calculations can be based on tests derived from a range of econometric estimators, including tobit, probit, and common panel data estimators.

⁵ See Feiveson (2002) for additional guidance on estimating the power of statistical tests using Stata.

⁶ There is not a universally employed method of standardization. Common alternatives include Glass’s Δ and Hedge’s *g*.

⁷ To be clear, if the standard deviation of the outcome variable is 2.5, a 0.10 standard deviation effect size refers to an unstandardized treatment effect of 0.25.

⁸ We have compiled a table of standardized effect sizes reported in experimental economics papers that present behavioral insights which can inform agri-environmental research and program design. The table will be periodically updated to reflect new research and can be found at <https://osf.io/cf259/>.

modify their research plan accordingly. Adequately powered experiments typically require larger sample sizes per treatment arm, resulting in costlier studies or the need to examine the effect of fewer treatments. Furthermore, authors must be transparent about which hypotheses were part of the original design and which hypotheses were generated later, after seeing the data, and thus should be seen more as *suggestive* than definitive.

Second, funding agencies must accept that more resources are needed to estimate treatment effects sufficiently, and proposals should be evaluated with this in mind. This expectation would appear to be at odds with common practice: the competitive nature of grant funding creates an incentive for researchers to propose projects that investigate numerous research questions while keeping costs low. Reviewers can support high quality research by requiring power analyses in proposals and having realistic expectations about the resources needed to conduct a robust study. Funding bodies may consider adopting policies that require submission of power analyses prior to releasing funds (similar to requirements to obtain approval to work with human subjects from Institutional Review Boards).⁹ Policies would need to be designed with care to limit attempts to game the system—e.g., positing very large minimum detectable treatment effects can increase the reported power but does not lead to a well-designed, sufficiently powered study.

Third, editors and reviewers must resist the temptation to discount small treatment effects in favor of larger, more exciting results. In economics and related disciplines, we have grown accustomed to reading papers that report large effect sizes. Instead, empirical manuscripts should be judged on the importance of the research question and the quality of the design (including power), not on the magnitude or the statistical significance of the estimates.

2.2 Methodological Challenge #2: Multiple Hypothesis Testing

During the conference, problems associated with multiple hypothesis testing (MHT) arose in three contexts: (1) testing the effects multiple treatments have on a single outcome; (2) testing the effects of a treatment on multiple outcomes; and (3) testing for heterogeneous treatment effects (subgroup effects). Testing multiple hypotheses within a single study is common in the agri-environmental literature, as well as the broader economics literature. For example, in a review of 34 articles using field experiments published in top economics journals, Fink et al. (2014) found that 76% of articles contained subgroup analysis, and 29% reported estimated treatment effects for ten or more subgroups. Testing multiple hypotheses is not inherently wrong; indeed, researchers are often rewarded for publishing many, meaningful and statistically significant results and reviewers often ask about treatment-effect heterogeneity and alternative model specifications. There are nevertheless two pitfalls associated with this practice.

First, given the rewards associated with having statistically significant results (e.g., having a paper accepted), researchers are tempted to seek out statistically significant results and discuss them loudly. However, this incentive can undermine the scientific process, particularly if researchers test multiple hypotheses and report *only* the results that are statistically significant (Simmons et al. 2011). Olken (2015) offers an illustrative example of

⁹ Power analyses are already an important component of proposals to fund clinical research, and an inadequate description of power calculations is considered to be a major issue during the review process (Inouye and Fiellin 2005).

how a deceptive researcher seeking only significant results could simply continue testing hypotheses until significant variables are discovered—regardless of whether the estimated effects reflect true effects. Even without deceptive intentions, well-meaning researchers can fall into this trap. We sensibly advocate that researchers report all results, even those from “failed” interventions.

Second, as the number of hypotheses tested increases, the chance increases for a researcher to make a false discovery; that is, to find a statistically significant effect even if a true effect does not exist (e.g., commit a Type I error) (Miguel et al. 2012). This is known as the multiple hypothesis testing (MHT) problem, which is recognized as a concern both in the broader social science literature and within experimental economics (Simmons et al. 2011; Olken 2015; List et al. 2016). In experimental economics, it is common for authors to test multiple hypotheses, but few studies undertake any corrections for MHT.

One approach to account for MHT is to limit the false discovery rate (FDR)—the expected proportion of false positives among the rejected hypotheses.¹⁰ To be clear, if a study has an FDR of 0.05 this means that, on average, only 5% of the rejected hypotheses represent *false* rejections. A common approach for controlling the FDR is the Benjamini and Hochberg procedure when the test statistics are independent (Benjamini and Hochberg 1995) or positively dependent (Benjamini and Yekutieli 2001). Among the alternatives under other forms of dependence include procedures proposed in Benjamini and Yekutieli (2001), Yekutieli (2008), and Xie et al. (2011). In brief, these procedures involve the use of alternative rejection rules (e.g., alternatives to “reject if $p < 0.05$ ”) that are conditional on the desired FDR, the number of hypotheses, and in most cases, the set of p-values obtained from the individual tests. For experimental economics studies of agri-environmental issues, we recommend using an FDR controlling procedure such that the FDR is 0.05 for “costly” treatments (e.g., testing incentives or dramatic changes to program structure), particularly when replications are unlikely. For “modest cost” treatments that are more likely to be replicated (e.g., initiatives that require data processing or collection, or require time or financial investments by producers), we recommend an FDR of 0.10. For “inexpensive” treatments (e.g., testing the effect of framing or information or lab experiments involving students), we recommend 0.20.

2.3 Methodological Challenge #3: Interpretation Issues

Controlled experimentation allows researchers to identify causal relationships, which is especially important when estimates are used to guide policy and program development (List and Price 2016). However, during the conference, three issues related to interpretation of the experimental results arose. First, randomization of treatments does not necessarily identify the underlying channels (mechanisms) through which a particular treatment works. To be clear, experimental designs *can* identify treatment effects, but they do not necessarily identify *why* treatment effects arise. For instance, a letter encouraging farmers to conserve water may motivate some to do so out of guilt, obligation, goodwill, a re-assessment

¹⁰ Another approach is to control the familywise error rate (FWER)—the probability of falsely rejecting even one hypothesis (i.e., the probability of at least one Type I error). List et al. (2019) present an approach to control the FWER that the authors assert leads to gains in power over Bonferroni-type procedures.

of actual water needs, and so on.¹¹ Identifying mechanisms requires additional assumptions, and is often achieved in experiments through combinations of theory and indirect tests (e.g., Ferraro and Miranda 2013). For example, if the aforementioned letter to farmers induces conservation by providing privately valuable information, we would expect farmers to be willing to pay a positive amount to receive another letter with updated information, but not if the letter operates through guilt. Therefore, although experiments can give credence to a particular mechanism's ability to affect behavior, researchers should be careful not to oversell statistically significant treatment effect estimates as proof that particular mechanisms "work".¹²

Second, in general, non-experimental variables included in the analysis of treatment effects cannot be interpreted causally. Through random treatment assignment, researchers can eliminate systematic correlation between unobservable characteristics and treatments. As a result, parameter estimates for variables that are assigned through randomization *can* be interpreted causally. Other covariates can, of course, be included in a regression analysis. Covariates are included for many reasons, such as to control for unintended differences in treatment samples, explore sources of heterogeneity of treatment effects or the outcome measure, and to increase the precision of the estimated treatment effects. However, it is important to articulate which estimates are identified through randomization and which are not. Estimated coefficients on variables that are not assigned randomly, such as socio-demographic characteristics of the experiment participants, responses to attitudinal questions, and elicited risk and time preferences, are more difficult to interpret because of classic econometric issues. If researchers want to make causal interpretations, we recommend that, as in observational studies, they employ careful identification strategies (e.g., structural modelling, instrumental variables estimation).

A third issue is the external validity of experimental results. For instance, can the results of a particular experiment be directly applied to a policy or program—e.g., are the results of the study directly applicable to the "real world"? Further, are certain subject pools more policy-relevant than others—e.g., does the behavior of students accurately reflect decisions farmers would make? Researchers often use laboratory experiments with student subject pools to test economic theories that are relevant to agri-environmental challenges (e.g., public good contributions, management of environmental externalities), but there is increasing interest in using field experiments with professional subject pools (i.e., farmers) in an effort to enhance external validity (Messer et al. 2014; Higgins et al. 2017).

In their paper in this special issue, Cason and Wu discuss the merits of student versus professional subject pools, and they address issues about the external validity of laboratory versus field experiments. Decisions about what type of experiment to run and which subjects to use are affected by research costs, control, replicability, and the ability to adequately answer the research questions. Importantly, the authors emphasize that one subject pool does not generate results that are more externally valid than another—the degree of external validity depends on the type of research question being addressed. They note that laboratory experiments with students "can enhance external validity because the experimenter can manipulate numerous variables and factors to put stress on the theory and

¹¹ If one can identify the various channels through which an intervention works, it may be possible to design an experiment that turns various channels on and off. See, for example, Ferraro and Hanauer (2014).

¹² For an overall discussion on the usefulness and limitations of randomized controlled trials, see Deaton and Cartwright (2018). For discussion about the need to further develop theoretical frameworks that can be used to generate testable hypotheses, see Muthukrishna and Henrich (2019).

determine how sensitive the predictions of the theory are to context” (Cason and Wu 2018). However, student responses to the modification of a specific program might not be as meaningful. For that type of research, a field experiment with farmers would generate more policy-relevant results. Researchers should recognize, however, that the external validity of field experiments is limited by the setting and characteristics of the participant sample. For example, results from a field experiment with farmers in the Northeastern United States will not necessarily generalize to the behavior of farmers in the Midwest or farmers in other countries. In general, Cason and Wu argue that laboratory experiments with student subjects are preferred when testing economic theory, whereas field experiments with professionals are preferred when research questions pertain to a specific policy or program and when it is important to measure preferences or characteristics of that population.

2.4 Methodological Challenge #4: Econometric Methods

To analyze the results of a controlled experiment, appropriate modes of statistical inference are critical. During the CBEAR-MAAP conference, we noticed two classes of issues related to statistical inference. The first class of issues involve what, unfortunately, are common mistakes when undertaking hypothesis tests. As one example, we noticed that authors were comparing treatment effects to a null hypothesis that the effect equaled zero and then drawing conclusions about the *relative* performance of different treatment arms. For instance, if Treatment A was statistically significant and Treatment B was not, one might mistakenly declare that Treatment A was more effective than Treatment B. Drawing that conclusion, however, requires a test of the null hypothesis that the two treatment effects are equal. As a second example, there were misgivings about the appropriate use (and form) of cluster-robust standard errors when analyzing repeated interventions. For instance, in one study the authors clustered the errors at the participant-level, although participants interacted in groups. In a second, the number of clusters was small, and the cluster-robust variance–covariance estimator is biased in this case. We refer the interested reader to Cameron and Miller’s (2015) extensive guide for practitioners on this topic.

A second class of issues related to the use of econometric identification strategies that are commonly used in the analysis of observational data in which treatments are not randomized. Such strategies included the use of difference-in-differences estimators and models with individual fixed effects. These approaches were presumably being used to control for unobservable individual heterogeneity. However, the researchers’ experimental design (i.e., randomization) should have already mitigated the correlation between unobservables and the treatment variables. Although applying these econometric methods to experimental data does not lead to invalid tests statistics, it adds unnecessary, additional structure to the estimation, which decreases the potential power of statistical tests by restricting the variation in the data used to identify treatment effects.

2.5 Methodological Advances Moving Forward

The methodological challenges described above can be addressed by investing more time in research planning and the experimental design stages of a project. First, we recommend that researchers publicly publish a pre-analysis plan in advance of data collection. Pre-analysis plans describe the research study, including the experimental design (if applicable) and sample sizes, and they specify how the resulting data will be analyzed, including the variables that will be collected, data cleaning procedures, and estimation strategies (Olken

2015). As a result, a pre-analysis plan serves as a commitment device, and helps circumvent several of the issues that we discussed above, such as underpowered designs. Moreover, a pre-analysis plan makes it more likely that researchers will be clearer when reporting results, in terms of specifying which results were related to the original aims of the experimental design and which were discovered during post-experiment exploratory analysis. To be clear, we are not suggesting that exploring the data and checking for interesting patterns is necessarily a devious practice. Indeed, this process often generates new hypotheses that can be tested in future work. We only encourage being clear about the nature of the patterns discovered in the research process. Authors should consider splitting the discussion of their results in a paper between the analysis originally outlined in the pre-analysis plan and the other exploratory analysis that was done after the data were collected. This type of approach would still enable exploratory analysis and the new discoveries that can arise, while also being transparent to the reader that these findings were not originally predicted in the initial pre-analysis plan.

Pre-analysis plans are required in the medical field for pre-clinical trials, including trials for pharmaceutical approval by the U.S. Food and Drug Administration. The American Economic Association sponsors a registry for randomized trials, and researchers are encouraged to register their trial and post their pre-analysis plan prior to data collection (www.socialscienceregistry.org/).¹³ The Open Science Framework also maintains a public registry where researchers can post pre-analysis plans for any type of study (<https://osf.io/registries/>).

Second, we recommend authors report standardized effect sizes as part of their results. With standardized effect sizes, readers can compare the magnitudes of estimated treatment effects across different treatments and outcomes. Researchers can refer to these effect sizes when priors are needed for their own power analyses. However, as noted in the section on underpowered designs, we caution researchers to be weary of determining their experiment design based on only large effect size estimates, as the resulting power analysis may suggest sample sizes that are too small to detect the true (smaller) effects.

Third, we recommend that economists join other disciplines, which make greater use of formal meta-analyses and study replications. Meta-analyses are a powerful, yet underutilized tool for synthesizing the experimental literature by pooling results of multiple studies. This pooling increases the precision of the overall treatment effect, as well as treatment effects conditional on moderating variables, including study design attributes. However, publication bias often limits the amount of information that is available for the meta-analysis. Ideally, we would have access to all of the studies, whether published or not, analyzing a particular behavior in order to determine the true effect size. When only certain types of results are published, this limits our ability to conduct meaningful meta-analysis. As long as a study is designed well, we should advocate for a greater willingness to publish null results, counterintuitive results, and replication studies. Replication studies have traditionally been difficult to publish in economics journals. Part of the stated focus of the recently created *Journal of the Economic Science Association* is to publish “article types that are important yet under-represented in the experimental literature”, including replications. We encourage other journals to be more willing to publish replication studies.

¹³ Currently, the AEA registry is set up primarily for the registration of RCTs, but we recommend that registries be updated to accommodate pre-analysis plans for both laboratory and field experiments.

3 Applying Behavioral Economics Insights to the Design of Agri-Environmental Programs

Over the past 30 years, behavioral scientists have both challenged and supported traditional economics assumptions about how individuals make decisions. In challenging these assumptions, they have demonstrated that cognitive and social factors once considered to be of second-order importance by economists can significantly affect people's decisions (Kahneman 2003; Leiser and Azar 2008; Kesternich et al. 2017). Moreover, these factors—such as the simplification and framing of information, the use of social norms and comparisons, and changes to the default choice—can often be easily and inexpensively modified in social programs (OECD 2017). These modifications, frequently called “nudges,” can alter decisions in predictable and cost-effective ways.¹⁴

By incorporating nudges and other behavioral insights in the design of agri-environmental programs and policies, administrators may be able to cost-effectively expand the impact of these programs (Higgins et al. 2017). However, most of the behavioral economics research related to program and policy design focuses on non-environmental issues, such as encouraging finance-friendly, health-friendly, education-friendly, or charity-friendly behaviors. Although there are behavioral and experimental economics studies with an agricultural or an environmental focus,¹⁵ few focus specifically on agri-environmental issues.

The absence of studies specifically testing the impact of behavioral nudges in agri-environmental policy settings is a problem because of two characteristics unique to the agri-environmental context. First, agri-environmental programs aim to affect long-term voluntary behaviors of agricultural producers who are often acting in competitive markets. Second, the long-term decisions being targeted are ones regarding the production of impure public goods. The broader behavioral science literature focuses on contexts that lack these two characteristics, and thus it is not clear whether the insights from that literature can be directly applied to the design of agri-environmental programs and policies.

A review that focuses specifically on the agri-environmental context is critical in order to take stock of what we know, but just as importantly, to highlight where research investments might best be directed to design efficient programs and policies. To classify behavioral insights that are relevant to agri-environmental programs, we use a framework we label Ag-E MINDSPACE, which is an extension of the MINDSPACE framework developed by Dolan et al. (2012).

¹⁴ Behavioral nudges were popularized by Thaler and Sunstein (2008) in their book, *Nudge*, which presents a behavioral economics toolkit for designing more-effective private and government programs and policies. They define a nudge as “any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates.” (Thaler and Sunstein 2008, p. 6).

¹⁵ See Mason and Phillips (1997), Messer et al. (2007), Shogren and Taylor (2008), Kotani et al. (2010), Messer and Murphy (2010), Shogren et al. (2010), Gsottbauer and van den Bergh (2011), Osbaldiston and Schott (2012), Friesen and Gangadharan (2013), Schultz (2014), List and Price (2016), Delaney and Jacobson (2016), Hobbs and Mooney (2016), Ferraro et al. (2017), Reddy et al. (2017), Zarghamee et al. (2017), Brent et al. (2017) and Kesternich et al. (2017).

Table 1 The MINDSPACE framework for behavioral change (from Dolan et al. 2012)

Cue	Behavior
Messenger	We are heavily influenced by who communicates information to us
Incentives	Our responses to incentives are shaped by predictable mental shortcuts such as strongly avoiding losses
Norms	We are strongly influenced by what others do
Defaults	We “go with the flow” of pre-set options
Salience	Our attention is drawn to what is novel and seems relevant to us
Priming	Our acts are often influenced by sub-conscious cues
Affect	Our emotional associations can powerfully shape our actions
Commitment	We seek to be consistent with our public promises, and reciprocate acts
Ego	We act in ways that make us feel better about ourselves

3.1 Ag-E MINDSPACE: Categorizing Behavioral Insights

Ag-E MINDSPACE comprises categories of nudges related to messengers, incentives, norms, defaults, salience, priming, affect, commitment, and ego, all of which can influence the behavior of agricultural producers (see Table 1).¹⁶ To develop a practical toolkit based on credible causal evidence, we established a set of criteria by which to select studies for review. In order to be included in this review, the published study had to¹⁷:

1. test the effect of one or more behavioral nudge(s) on producer decision-making.
2. be motivated by an agri-environmental challenge or program.
3. have economic content or consequences.
4. employ an experimental design in which the measured behavioral outcomes are revealed behaviors with salient costs and benefits (rather than stated preferences).

In Table 2, we assign each of the studies we identified to one or more of the nine Ag-E MINDSPACE categories. Some studies fit into more than one Ag-E MINDSPACE category because the nudges analyzed address more than one behavioral insight. In this section, we describe the key behavioral insights gained from studies in each category.

3.1.1 Messenger

Responses to an intervention can be strongly influenced (positively or negatively) by the messenger who delivers information. Messengers may be particularly important when addressing controversial issues such as climate change or unpopular government programs. It would be beneficial for such programs to test various messages and messengers using

¹⁶ Although consumer preferences and behavior can drive change in agriculture, we include only studies analyzing producer behavior. This is consistent with the aim of agri-environmental programs, which is to achieve permanent changes in how producers manage impure public goods.

¹⁷ We only review papers that are published in peer-reviewed journals, but we acknowledge that there is a growing body of experimental literature in the agri-environmental domain, and many of these papers are in the review process or in working paper form. We do not believe, however, that including this body of in-progress work would change our overall conclusions.

administrative experiments to identify what type of information and which information sources (messengers) will most effectively break down barriers to increase participation and change targeted producer behaviors. We are aware of only one study that has been published on the effect of messengers in an agri-environmental context (Butler et al. 2019) and another that is a working paper (Griesinger et al. 2017).

3.1.2 Incentives

Many federal, state, and local agri-environmental programs use monetary incentives to motivate voluntary changes in land management practices. Although economists have long studied the effects of incentives in agri-environmental contexts, they typically have ignored how the incentives are framed, how they are complemented by behavioral nudges, or how other nonpecuniary aspects of program structure and delivery can affect behavior. Dolan et al. (2012) offer useful insights about these oft-ignored aspects of incentives: (1) reference points matter; (2) losses loom larger than gains (loss aversion); (3) small probabilities are overweighted; (4) money is allocated mentally to discrete accounts (mental accounting); and (5) choices consistently reflect living for today at the expense of tomorrow (present bias).

Research on point (5) has shown that farmers, like other economic agents, prefer immediate payoffs relative to larger payoffs in the future (Duflo et al. 2011; Duquette et al. 2012).¹⁸ Furthermore, Duquette et al. (2012) find that farmers who were considered “late adopters” of agricultural BMPs have discount rates that are higher than average. Such findings have implications for the design of agri-environmental programs in which it is often difficult to engage farmers who are resistant to try new BMPs. Although there exists a substantial body of agri-environmental experimental economics literature on the *presence* of incentives,¹⁹ far fewer studies investigate the importance of how producer behavior changes in response to *how* incentives are presented (e.g., how incentives are framed).

3.1.3 Norms

In some situations, individuals take their cues from what others do; therefore, providing information on social norms can strongly influence a person’s behavior (Burke and Young 2011). Research has demonstrated the positive impact of social norms in improving environmental decision making (e.g., motivating conservation behavior) (Ferraro et al. 2011; Allcott 2011; Ferraro and Price 2013; Bernedo and Ferraro 2017), but we know of only one study that tests the effects of providing social norm information to agricultural producers. In an administrative experiment, Wallander et al. (2017) study the effects of sending messages to farmers considering enrollment in the Conservation Reserve Program

¹⁸ Duflo et al. (2011) analyze farmer investments in fertilizer and is, therefore, not categorized as an Ag-E paper.

¹⁹ A large portion of this literature tests the outcomes of various tax and subsidy mechanisms to reduce nonpoint source (NPS) pollution (Alpizar et al. 2004; Poe et al. 2004; Spraggon 2004, 2013; Cochard et al. 2005; Vossler et al. 2006; Suter et al. 2008; Spraggon and Oxoby 2010; Cason 2010; Cason and Gangadharan 2013; Suter and Vossler 2014; Miao et al. 2016; Palm-Forster et al. 2017), improve extraction of ground water for irrigation (Gardner, Moore, and Walker 1997; Suter et al. 2012, 2018; Li et al. 2014; Liu et al. 2014), and incentivize land conservation and ecosystem service provision (Parkhurst et al. 2002; Cason and Gangadharan 2004; Parkhurst and Shogren 2007, 2008; Arnold, Duke, and Messer 2013; Banerjee et al. 2014, 2015, 2017; Fooks et al. 2015, 2016; Messer et al. 2017; Duke et al. 2017; Banerjee 2018; Reeling et al. 2018).

Table 2 Ag-E MINDSPACE: evidence of agri-environmental behavioral insights^a

MINDSPACE category	Experimental economics studies on producer behavior
Messenger	Butler et al. (2019) ^{Lab}
Incentives ^b	Duquette et al. (2012) ^{Adm} Li et al. (2014) ^{Lab}
Norms	Banerjee et al. (2014) ^{Lab} Banerjee (2018) ^{Lab} Wallander et al. (2017) ^{Adm}
<i>Defaults</i>	
Salience	Banerjee et al. (2015) ^{Lab} Cason et al. (2003) ^{Lab} Higgins et al. (2017) ^{Adm} Li et al. (2014) ^{Lab} Messer et al. (2017) ^{Lab} Wallander et al. (2017) ^{Adm}
Priming	Cason et al. (2003) ^{Lab} Czap et al. (2013) ^{Lab} Higgins et al. (2017) ^{Adm} Wallander et al. (2017) ^{Adm}
Affect	Czap et al. (2013) ^{Lab} Messer and Borchers (2015) ^{FFE}
Commitment	
Ego	

Lab laboratory experiment, *Arte* artefactual experiment, *FFE* framed field experiment, *Adm* administrative experiment [see Harrison and List (2004), Messer et al. (2014), and Higgins et al. (2017) for definitions of types of experiments]

^aA table of standardized effect sizes related to this literature will be periodically updated to reflect the growing evidence base and can be found at <https://osf.io/cf259/>

^bEconomists have long studied the effects of incentives in agri-environmental contexts; however, this reference focuses on less studied features of incentives, such as how they are framed or complementarities between incentives and behavioral nudges

(CRP). The authors are unable to detect any difference in sign-up rates between farmers who received a simple reminder about the enrollment period and farmers who received the reminder augmented with social comparisons or injunctive norms. Group interaction, such as communication and public voting, has been shown to be an important means of developing social norms.²⁰

²⁰ See Ostrom (2000) for more discussion on the evolution of rules and social norms. There is also broad literature on the role that communication and voting have on improving the performance of groups in public good and common pool resource settings (Messer et al. 2017, 2008); however, these studies have not focused on agri-environmental decision-making.

3.1.4 Defaults

Default options (also known as a status-quo options) often serve as strong reference points and are frequently passively chosen over alternatives (Dolan et al. 2012).²¹ Defaults have been lauded in behavioral design because they can influence decisions without removing the individual's ability to choose. Defaults exist in agri-environmental programs, but few programs are using them strategically. In other environmental contexts such as enrollment in electricity programs, defaults for automatic enrollment provide an unusually simple and nearly cost-free way to increase program participation (Fowlie et al. 2017). Defaults may be especially useful when choices are complex. When faced with a choice that is difficult to analyze, participants may be more likely to accept the default option. Although we are aware of research in progress that studies the effect of defaults in agri-environmental cost-share programs, we do not know of any published studies on the topic.

3.1.5 Saliency

Dolan et al. (2012) note that people's choices are influenced by which parts of the decision draw their attention, which are typically facets of the decision that are salient and easily understood. The influence of saliency points to the need for clear, concise, nontechnical explanations in program materials and communications. Given its broad subjective scope, saliency tends to overlap with several other types of nudges, such as norms, priming, and affect, which require disclosure and dissemination of information. However, providing information that is more salient can potentially have adverse effects, such as increasing rent-seeking. In laboratory experiments, Cason et al. (2003), Banerjee et al. (2015), and Messer et al. (2017) observe rent-seeking behavior in conservation auctions when producers are given salient information that allowed them to identify the environmental quality of their lands and the environmental benefits of the auction, respectively. However, if the goal is to increase the amount of contiguous land conserved by an agglomeration bonus, more information can be better. In a laboratory experiment, Banerjee et al. (2014) show that spatial coordination and efficiency improves when potential participants are given salient information about their neighbors' behavior.

3.1.6 Priming

Priming refers to influencing decisions through subconscious cues, like words, sights, and sounds (Dolan et al. 2012). The fifth case study described in Higgins et al. (2017) demonstrates how priming can be used to encourage certain behavior. The administrative experiment was motivated by declining participation in USDA's Farm Service Agency (FSA) county committee elections. The researchers tested the ability of priming nudges to motivate agricultural producers to participate in the 2015 elections. They tested the effects of postcard reminders mailed a week before and after the election deadline, and presentation of the candidates' names on the outside of the mailed ballot. They find that these two simple nudges increased participation. In a laboratory experiment, Czap et al. (2013) find that priming messages included in the experiment instructions increase participants' conservation behavior. However, other studies have shown that this provision of information about what an agri-environmental program

²¹ As a frequently cited example of the power of defaults, organ donor rates are much higher in countries where the default option is that everyone is an organ donor (Johnson and Goldstein 2003).

seeks to target for investment can lead to rent seeking behavior by landowners who suspect that they are being targeted (Cason et al. 2003; Fooks et al. 2016).

3.1.7 Affect

Affect describes cases in which people's emotional responses to words, images, and events change the way that they view and value various options. Those changes can be short-lived or persist for longer periods of time (Dolan et al. 2012). Such emotional nudges can be used by agri-environmental practitioners when designing the framing and content of proposals, interfaces, and specific messages to encourage consumers and agricultural producers to connect their actions to the external impacts they create. Emotional nudges cost little to implement, yet the resulting emotional responses may increase the value of the program to potential participants and improve program outcomes. Czap et al. (2013) conduct a laboratory experiment that emulates a water pollution problem and test three framings when presenting instructions regarding how upstream producers' actions affected outcomes in the social-ecological system: neutral (no context); empathy (empathetic to the downstream water user); and self-interest (profit maximization). They find that the empathy frame increases pollution abatement. In a framed field experiment about land conservation, Messer and Borchers (2015) find that the credible threat of destructing wine (representing land) leads to a significant increase in the preference for selecting (and thus protecting) rare and expensive wine (70.1% versus 29.9% with no threat).

3.1.8 Commitment

Pledges, oaths, and commitments are an integral part of our society and are required by many professions, including medical doctors and elected government officials. Recent experimental research finds that oath-taking can improve coordination and lead to more efficient outcomes in strategic environments (Jacquemet et al. 2018). These findings have implications for the provision of environmental goods, which often requires coordination among numerous individuals. In agri-environmental settings, commitments can be used as an inexpensive nudge by asking producers and consumers to make voluntary conservation promises or pledges. In their meta-analysis of studies of pro-environmental commitments, Lokhorst et al. (2013) find that the studies produce mixed results, but commitment devices are effective when certain techniques are followed. The type of commitment required and the form in which it is provided (electronic or written, requested in person or by mail) can affect the ability of the commitment to produce results, and the presence of a referee or credible audit can enhance the effectiveness of the commitment device. Commitments can also be combined with other MINDSPACE nudges. Public displays of a commitment, such as signs that indicated that a farm is enrolled in a conservation program or is permanently protected from development, would provide an *ego* nudge as well, potentially motivating participants to make good on their pledges. While some studies have begun to study this issue (i.e. Griesinger et al. 2017), we know of no published agri-environmental experiments that test the impact of commitments.

3.1.9 Ego

In the MINDSPACE framework, *ego* refers to the human desire to have a consistent identity and/or positive self-image. Norms play a part in ego—we (mostly) strive to avoid signals of repugnance from others, shame, and conflict and, therefore, often adhere to cultural norms (see

discussion in Dolan et al. 2012). Agricultural producers may be motivated by a desire to be seen as protecting the environment. *Ego* can also overlap with *commitment* if there is a feedback mechanism such as a referee or monitor that audits an individual's commitment. Agri-environmental stewardship certification programs and stewardship awards may be powerful tools to motivate behavioral change because of the recognition that they offer to farmers who are environmental stewards. However, it is unclear which of the benefits provided by these programs is the strongest incentive to encourage stewardship. For example, recognition from stewardship certification programs—like the Michigan Agriculture Environmental Assurance Program (MAEAP)—is often coupled with benefits such as insurance discounts and legal protections that motivate participation (Stuart et al. 2014). Controlled artefactual or framed field experiments could be used to identify how nudges that affect one's ego impact behavior.²²

4 Conclusion and Discussion

There is growing interest in using behavioral and experimental economics research to inform agri-environmental policy and program design, but the associated literature that studies the behavior of agricultural producers is limited. New research filling this gap was presented at the 2017 Conference on Behavioral and Experimental Agri-environmental Research: Methodological Advancements and Applications to Policy (CBEAR-MAAP). Selected papers from the conference are included in this special issue in revised form. In this article, we highlighted methodological challenges and approaches to overcome them, and we reviewed the small, but growing literature on behavioral insights that can inform the design of agri-environmental programs and policies. We conclude by summarizing our recommendations and suggesting ways to prioritize future policy-relevant research.

4.1 Summary of Recommendations

Four key methodological challenges that arose during the conference were: (1) underpowered designs; (2) multiple hypothesis testing; (3) interpretation of results; and (4) econometric methods. To overcome these challenges, we have ten recommendations:

1. Dedicate sufficient time and resources in the experimental design stage to conduct power analyses and publish a pre-analysis plan on a public experiment registry that describes each treatment, the experimental design, the required sample size, testable hypotheses, and statistical procedures.²³
2. Design the experiment to detect an effect size that is smaller than expected. Effect sizes estimated with a sufficiently powered design are likely to be considerably smaller than those reported in related studies, due to factors such as publication bias and underpowered designs. Based on the standardized effect sizes we report in this literature, designs that are sufficiently powered to detect at least a 0.10 standard deviation effect size, or even smaller, are often advisable.

²² See Harrison and List (2004), Messer, Duke, and Lynch (2014), and Higgins et al. (2017) for definitions of types of experiments.

²³ We recommend publishing pre-analysis plans on a public experiment registry, like those maintained by the American Economic Association and the Open Science Framework.

3. Use appropriate econometric methods for experimental data. Avoid using difference-in-differences estimators and models with individual fixed effects to estimate the effect of randomized treatments, which are inefficient approaches that unnecessarily decrease power.
4. Avoid causally interpreting the effects of non-experimental variables unless, as in careful observational studies, an appropriate identification strategy is employed for those variables.
5. Account for multiple hypothesis testing (MHT) using approaches that limit the false discovery rate, such as the Benjamini and Hochberg procedure. We recommend controlling for a false discovery rate of 0.05 for “costly” treatments, 0.10 for “modest cost” treatments that are more likely to be replicated, and 0.20 for “inexpensive” treatments, such as those with student participants.
6. Report all of the results for analyses outlined in the pre-analysis plan, even those associated with null findings.
7. Report standardized effect sizes so that readers can compare the magnitudes of estimated treatment effects across different treatments and outcomes.
8. Journal editors and manuscript reviewers should assess the quality of the research instead of the magnitude of the results when deciding which papers warrant publication.
9. Publication of formal meta-analyses and replication studies should be encouraged.
10. Funding agencies and proposal review panels should support high quality research by requiring power analyses and pre-analysis plans in proposals for funding, and having realistic expectations about the resources needed to conduct a robust study.

4.2 Prioritizing Behavioral and Experimental Agri-Environmental Research

Researchers and practitioners have emphasized the importance of designing more cost-effective agri-environmental programs and policies in order to generate the most value from limited budgets. However, although a significant amount of research has been conducted on the design and performance of various market and incentive-based mechanisms, fewer studies have analyzed how insights from the broader behavioral sciences can be used to improve policy and program design. We find that, unlike with other policy domains in which one can find dozens of relevant behavioral studies, the agri-environmental domain is characterized by a paucity of relevant studies that can guide practitioners. Practitioners are thus forced to (1) assume that results from other domains, which are often derived from consumer decisions in private good settings, can be applied to agri-environmental contexts, which often involve producer decisions that affect both private and public goods, (2) collaborate with researchers to replicate and extend the insights from other domains to important agri-environmental contexts, or (3) ignore the potential benefits that could arise from applying these approaches.

Based on the sparse experimental economics literature on agri-environmental issues, in particular as it pertains to “nudges”, there is tremendous opportunity for new work to fill in the gaps. But what should researchers tackle first? In making decisions about research priorities, we recommend weighing three factors: (1) the cost of conducting a well-designed experiment to test a particular intervention; (2) the expected net social benefits of the intervention being tested; and (3) the uncertainty surrounding the expected net benefits of the intervention. When referring to costs of conducting an experiment, we implicitly include

both financial and ethical costs. When referring to net social benefits, we include the potential benefits of the intervention in real-world programs, as well as the potential benefits from a better understanding of the underlying behavioral mechanisms connected to agri-environmental outcomes, which have been debated for decades (e.g., do producers fail to adopt conservation practices because of information and cognitive constraints or because the practices are simply not financially lucrative?).

Ample opportunities exist for applications characterized by low costs and high expected net benefits, but with enough uncertainty about impact levels to warrant additional research. Costly projects that investigate interventions with potentially high, but uncertain benefits are likely also high priority. In contrast, an intervention that costs little or nothing to implement within a program would not be a strong candidate for expensive experimental testing. Neither would a project with high costs and low expected benefits. Researchers could likely use project funds more wisely by analyzing other interventions with the potential to provide greater net benefits to society.

The majority of the current experimental agri-environmental literature consists of laboratory experiments, and artefactual and framed field experiments. This work has allowed researchers to test and further develop economic theories, and it has contributed to our understanding of how people respond to incentives. Moreover, experiments exercising greater control continue to be important for test bedding interventions prior to field implementation. Moving forward, we see great potential for taking more experiments into the field, including embedding experiments in agri-environmental programs. Further, we encourage more research that tests interventions with multiple subject pools in a variety of settings and contexts to enhance the external validity of experimental research. Program administrators are constantly trying out new ways to design programs and deliver information, but few of these changes are implemented in a controlled way that permits causal interpretation of outcome changes. Researchers and program administrators have overlapping interests that make them natural partners to test new interventions.

Acknowledgements The authors acknowledge financial support from the USDA Economic Research Service (ERS), the Center for Behavioral and Experimental Agri-environmental Research (CBEAR), and conference funding from the USDA National Institute of Food and Agriculture through the Agriculture and Food Research Initiative Foundational Program (NIFA-AFRI Grant No. 12234087). The authors appreciate feedback and helpful comments from Simanti Banerjee, Tim Cason, Lata Gangadharan, Jordan Suter, Tim Wojan, attendees at the 2018 Brown Bag Lunch Series on Behavioral Science in Agri-environmental Program Design hosted by USDA ERS, and participants of the 2018 Appalachian Experimental and Environmental Economics Workshop and the 2017 Conference for Behavioral and Experimental Agri-environmental Research: Methodological Advancements and Applications to Policy (CBEAR-MAAP). Dr. Janusch contributed to this article in his personal capacity. The views expressed are his own and do not necessarily represent the views of the California Energy Commission.

References

- Allcott H (2011) Social norms and energy conservation. *J Public Econ* 95:1082–1095. <https://doi.org/10.1016/j.jpubeco.2011.03.003>
- Alpízar F, Requate T, Schram A (2004) Collective versus random fining: an experimental study on controlling ambient pollution. *Environ Resour Econ* 29:231–252. <https://doi.org/10.1023/B:EARE.0000044608.66145.0c>
- Armitage P, McPherson CK, Rowe BC (1969) Repeated significance tests on accumulating data. *J R Stat Soc Ser Gen* 132:235–244. <https://doi.org/10.2307/2343787>
- Arnold MA, Duke JM, Messer KD (2013) Adverse selection in reverse auctions for ecosystem services. *Land Econ* 89:387–412

- Banerjee S (2018) Improving spatial coordination rates under the agglomeration bonus scheme: a laboratory experiment with a pecuniary and a non-pecuniary mechanism (NUDGE). *Am J Agric Econ* 100:172–197. <https://doi.org/10.1093/ajae/aa066>
- Banerjee S, de Vries FP, Hanley N, van Soest DP (2014) The impact of information provision on agglomeration bonus performance: an experimental study on local networks. *Am J Agric Econ* 96:1009–1029. <https://doi.org/10.1093/ajae/aau048>
- Banerjee S, KwASNICA AM, Shortle JS (2015) Information and auction performance: a laboratory study of conservation auctions for spatially contiguous land management. *Environ Resour Econ* 61:409–431. <https://doi.org/10.1007/s10640-014-9798-4>
- Banerjee S, Cason TN, de Vries FP, Hanley N (2017) Transaction costs, communication and spatial coordination in payment for ecosystem services schemes. *J Environ Econ Manag* 83:68–89. <https://doi.org/10.1016/j.jeem.2016.12.005>
- Bellemare C, Bissonnette L, Kröger S (2016) Simulating power of economic experiments: the powerBBK package. *J Econ Sci Assoc* 2:157–168. <https://doi.org/10.1007/s40881-016-0028-4>
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29:1165–1188. <https://doi.org/10.1214/aos/1013699998>
- Bernedo M, Ferraro PJ (2017) Behavioral economics and climate change adaptation: insights from experimental economics on the role of risk and time preferences. In: Botelho A (ed) *World scientific reference on natural resources and environmental policy in the era of global change: volume 4: experimental economics*. World Scientific, pp 151–177
- Brent DA, Friesen L, Gangadharan L, Leibbrandt A (2017) Behavioral insights from field experiments in environmental economics. *Int Rev Environ Resour Econ* 10:95–143. <https://doi.org/10.1561/101.00000084>
- Brown JP, Lambert DM, Wojan TR (2019) The effect of the conservation reserve program on rural economies: deriving a statistical verdict from a null finding. *Am J Agric Econ* 101:528–540. <https://doi.org/10.1093/ajae/aay046>
- Burke MA, Young HP (2011) Social norms. In: Benhabib J, Bisin A, Jackson MO (eds) *Handbook of social economics*. North-Holland, Amsterdam, pp 311–338
- Butler JM, Fooks JR, Messer KD, Palm-Forster LH (2019) Addressing social dilemmas with mascots, information, and graphics. *Econ Inq*. <https://doi.org/10.1111/ecin.12783>
- Button KS, Ioannidis JPA, Mokrysz C et al (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14:365–376. <https://doi.org/10.1038/nrn3475>
- Cameron AC, Miller DL (2015) A practitioner's guide to cluster-robust inference. *J Hum Resour* 50:317–372
- Cason TN (2010) What can laboratory experiments teach us about emissions permit market design? *Agric Resour Econ Rev* 39:151–161. <https://doi.org/10.1017/S1068280500007218>
- Cason TN, Gangadharan L (2004) Auction design for voluntary conservation programs. *Am J Agric Econ* 86:1211–1217. <https://doi.org/10.1111/j.0002-9092.2004.00666.x>
- Cason TN, Gangadharan L (2013) Empowering neighbors versus imposing regulations: an experimental analysis of pollution reduction schemes. *J Environ Econ Manag* 65:469–484. <https://doi.org/10.1016/j.jeem.2012.09.001>
- Cason TN, Wu SY (2018) Subject pools and deception in agricultural and resource economics experiments. *Environ Resour Econ*. <https://doi.org/10.1007/s10640-018-0289-x>
- Cason TN, Gangadharan L, Duke C (2003) A laboratory study of auctions for reducing non-point source pollution. *J Environ Econ Manag* 46:446–471. [https://doi.org/10.1016/S0095-0696\(03\)00026-3](https://doi.org/10.1016/S0095-0696(03)00026-3)
- Cherry TL, Kroll S, Shogren JF (eds) (2008) *Environmental economics, experimental methods*. Routledge, New York
- Cochard F, Willinger M, Xepapadeas A (2005) Efficiency of nonpoint source pollution instruments: an experimental study. *Environ Resour Econ* 30:393–422. <https://doi.org/10.1007/s10640-004-5986-y>
- Cohen J (1988) *Statistical power analysis for the behavioral sciences*, 2nd edn. Lawrence Erlbaum Associates, Hillsdale
- Czap NV, Czap HJ, Khachatryan M et al (2013) Smiley or Frowney: the effect of emotions and empathy framing in a downstream water pollution game. *Int J Econ Finance* 5:9. <https://doi.org/10.5539/ijef.v5n3p9>
- Deaton A, Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 210:2–21. <https://doi.org/10.1016/j.socscimed.2017.12.005>

- Delaney J, Jacobson S (2016) Payments or persuasion: common pool resource management with price and non-price measures. *Environ Resour Econ* 65:747–772. <https://doi.org/10.1007/s10640-015-9923-z>
- Dolan P, Hallsworth M, Halpern D et al (2012) Influencing behaviour: the mindspace way. *J Econ Psychol* 33:264–277. <https://doi.org/10.1016/j.joep.2011.10.009>
- Doucouliaogis C, Stanley TD (2013) Are all economic facts greatly exaggerated? Theory competition and selectivity. *J Econ Surv* 27:316–339
- Duflo E, Kremer M, Robinson J (2011) Nudging farmers to use fertilizer: theory and experimental evidence from Kenya. *Am Econ Rev* 101:2350–2390. <https://doi.org/10.1257/aer.101.6.2350>
- Duke JM, Messer KD, Lynch L, Li T (2017) The effect of information on discriminatory-price and uniform-price reverse auction efficiency: an experimental economics study of the purchase of ecosystem services. *Strateg Behav Environ* 7:41–71. <https://doi.org/10.1561/102.00000073>
- Duquette E, Higgins N, Horowitz J (2012) Farmer discount rates: experimental evidence. *Am J Agric Econ* 94:451–456. <https://doi.org/10.1093/ajae/aar067>
- Ehmke MD, Shogren JF (2009) Experimental methods for environment and development economics. *Environ Dev Econ* 14:419–456. <https://doi.org/10.1017/S1355770X08004592>
- Ellis PD (2010) The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results. Cambridge University Press, New York
- FAO (2016) FAOSTAT land domain. <http://www.fao.org/faostat/en/#data/RL>. Accessed 13 Mar 2019
- Feiveson AH (2002) Power by simulation. *Stata J* 2:107–124. <https://doi.org/10.1177/1536867X0200200201>
- Ferraro PJ, Hanauer MM (2014) Quantifying causal mechanisms to determine how protected areas affect poverty through changes in ecosystem services and infrastructure. *Proc Natl Acad Sci*. <https://doi.org/10.1073/pnas.1307712111>
- Ferraro PJ, Miranda JJ (2013) Heterogeneous treatment effects and mechanisms in information-based environmental policies: evidence from a large-scale field experiment. *Resour Energy Econ* 35:356–379. <https://doi.org/10.1016/j.reseneeco.2013.04.001>
- Ferraro PJ, Price MK (2013) Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Rev Econ Stat* 95:64–73. https://doi.org/10.1162/REST_a_00344
- Ferraro PJ, Miranda JJ, Price MK (2011) The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *Am Econ Rev* 101:318–322
- Ferraro PJ, Messer KD, Wu S (2017) Applying behavioral insights to improve water security. *Choices* 32:1–6
- Fink G, McConnell M, Vollmer S (2014) Testing for heterogeneous treatment effects in experimental data: false discovery risks and correction procedures. *J Dev Eff* 6:44–57. <https://doi.org/10.1080/19439342.2013.875054>
- Fooks JR, Messer KD, Duke JM (2015) Dynamic entry, reverse auctions, and the purchase of environmental services. *Land Econ* 91:57–75
- Fooks JR, Higgins N, Messer KD et al (2016) Conserving spatially explicit benefits in ecosystem service markets: experimental tests of network bonuses and spatial targeting. *Am J Agric Econ* 98:468–488. <https://doi.org/10.1093/ajae/aav061>
- Fowle M, Wolfram C, Spurlock CA et al (2017) Default effects and follow-on behavior: evidence from an electricity pricing program. National Bureau of Economic Research, Cambridge
- Friesen L, Gangadharan L (2013) Environmental markets: what do we learn from the lab? *J Econ Surv* 27:515–535. <https://doi.org/10.1111/joes.12021>
- Gardner R, Moore MR, Walker JM (1997) Governing a groundwater commons: a strategic and laboratory analysis of western water law. *Econ Inq* 35:218–234. <https://doi.org/10.1111/j.1465-7295.1997.tb01905.x>
- Gelman A, Carlin J (2014) Beyond power calculations: assessing type S (sign) and type M (magnitude) errors. *Perspect Psychol Sci* 9:641–651. <https://doi.org/10.1177/1745691614551642>
- Griesinger MR, Palm-Forster LH, Messer KD et al (2017) Stewardship signaling and the power of using social pressures to reduce nonpoint source pollution. Agricultural and Applied Economics Association, Chicago
- Gsottbauer E, van den Bergh JCJM (2011) Environmental policy theory given bounded rationality and other-regarding preferences. *Environ Resour Econ* 49:263–304. <https://doi.org/10.1007/s10640-010-9433-y>
- Harrison GW, List JA (2004) Field experiments. *J Econ Lit* 42:1009–1055
- Higgins N, Hellerstein D, Wallander S, Lynch L (2017) Economic experiments for policy analysis and program design: a guide for agricultural decisionmakers. US Department of Agriculture, Economic Research Service, Washington



- Hobbs JE, Mooney S (2016) Applications of behavioral and experimental economics to decision making in the agricultural, food, and resource sectors: an introduction. *Can J Agric Econ Can Agroekon* 64:593–597. <https://doi.org/10.1111/cjag.12117>
- Inouye SK, Fiellin DA (2005) An evidence-based guide to writing grant proposals for clinical research. *Ann Intern Med* 142:274–282. <https://doi.org/10.7326/0003-4819-142-4-200502150-00009>
- Ioannidis JPA, Stanley TD, Doucouliagos H (2017) The power of bias in economics research. *Econ J* 127:F236–F265. <https://doi.org/10.1111/econj.12461>
- Jacquemet N, Luchini S, Shogren JF, Zylbersztejn A (2018) Coordination with communication under oath. *Exp Econ* 21:627–649. <https://doi.org/10.1007/s10683-016-9508-x>
- Johnson EJ, Goldstein D (2003) Do defaults save lives? *Science* 302:1338–1339. <https://doi.org/10.1126/science.1091721>
- Kahneman D (2003) Maps of bounded rationality: psychology for behavioral economics. *Am Econ Rev* 93:1449–1475
- Kesternich M, Reif C, Rübhelke D (2017) Recent trends in behavioral environmental economics. *Environ Resour Econ* 67:403–411. <https://doi.org/10.1007/s10640-017-0162-3>
- Kotani K, Messer KD, Schulze WD (2010) Matching grants and charitable giving: why people sometimes provide a helping hand to fund environmental goods. *Agric Resour Econ Rev* 39:324–343
- Lakens D, Adolfs FG, Albers CJ et al (2018) Justify your alpha. *Nat Hum Behav* 2:168–171. <https://doi.org/10.1038/s41562-018-0311-x>
- Leiser D, Azar OH (2008) Behavioral economics and decision making: applying insights from psychology to understand how people make economic decisions. *J Econ Psychol* 29:613–618. <https://doi.org/10.1016/j.joep.2008.08.001>
- Li J, Michael HA, Duke JM et al (2014) Behavioral response to contamination risk information in a spatially explicit groundwater environment: experimental evidence. *Water Resour Res* 50:6390–6405. <https://doi.org/10.1002/2013WR015230>
- List JA, Price MK (2016) The use of field experiments in environmental and resource economics. *Rev Environ Econ Policy*. <https://doi.org/10.1093/reep/rew008>
- List JA, Sadoff S, Wagner M (2011) So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Exp Econ* 14:439–457. <https://doi.org/10.1007/s10683-011-9275-7>
- List JA, Shaikh AM, Xu Y (2016) Multiple hypothesis testing in experimental economics. National Bureau of Economic Research, Cambridge
- List JA, Shaikh A, Xu Y (2019) Multiple hypothesis testing in experimental economics. *Exp Econ*. <https://doi.org/10.1007/s10683-018-09597-5>
- Liu Z, Suter JF, Messer KD et al (2014) Strategic entry and externalities in groundwater resources: evidence from the lab. *Resour Energy Econ* 38:181–197. <https://doi.org/10.1016/j.reseneeco.2014.07.002>
- Lokhorst AM, Werner C, Staats H et al (2013) Commitment and behavior change: a meta-analysis and critical review of commitment-making strategies in environmental research. *Environ Behav* 45:3–34. <https://doi.org/10.1177/0013916511411477>
- Mason CF, Phillips OR (1997) Mitigating the tragedy of the commons through cooperation: an experimental evaluation. *J Environ Econ Manag* 34:148–172. <https://doi.org/10.1006/jeem.1997.1006>
- Messer KD, Borchers AM (2015) Choice for goods under threat of destruction. *Econ Lett* 135:137–140. <https://doi.org/10.1016/j.econlet.2015.07.026>
- Messer KD, Murphy JJ (2010) Special issue on experimental methods in environmental, natural resource, and agricultural economics. *Econ Rev* 39:iii–vi. <https://doi.org/10.1017/s106828050000719x>
- Messer KD, Kaiser HM, Poe GL (2007) Voluntary funding for generic advertising using a provision point mechanism: an experimental analysis of option assurance. *Appl Econ Perspect Policy* 29:612–631. <https://doi.org/10.1111/j.1467-9353.2007.00375.x>
- Messer KD, Duke JM, Lynch L (2014) Applying experiments to land economics: public information and auction efficiency in ecosystem service markets. In: Duke JM, Wu J (eds) *The Oxford handbook of land economics*. Oxford University Press, Oxford, pp 481–546
- Messer KD, Duke JM, Lynch L, Li T (2017) When does public information undermine the efficiency of reverse auctions for the purchase of ecosystem services? *Ecol Econ* 134:212–226. <https://doi.org/10.1016/j.ecolecon.2016.12.004>
- Miao H, Fooks JR, Guilfoos T et al (2016) The impact of information on behavior under an ambient-based policy for regulating nonpoint source pollution. *Water Resour Res* 52:3294–3308. <https://doi.org/10.1002/2015WR018142>
- Miguel E, Casey K, Glennerster R (2012) Reshaping institutions: evidence on aid impacts using a preanalysis plan. *Q J Econ* 127:1755–1812. <https://doi.org/10.1093/qje/qje027>

- Muthukrishna M, Henrich J (2019) A problem in theory. *Nat Hum Behav* 10:10. <https://doi.org/10.1038/s41562-018-0522-1>
- OECD (2017) Tackling environmental problems with the help of behavioural insights. Organisation for Economic Cooperation and Development, Paris. <https://doi.org/10.1787/9789264273887-en>
- Olken BA (2015) Promises and perils of pre-analysis plans. *J Econ Perspect* 29:61–80. <https://doi.org/10.1257/jep.29.3.61>
- Osbaldiston R, Schott JP (2012) Environmental sustainability and behavioral science: meta-analysis of proenvironmental behavior experiments. *Environ Behav* 44:257–299. <https://doi.org/10.1177/0013916511402673>
- Ostrom E (2000) Collective action and the evolution of social norms. *J Econ Perspect* 14:137–158. <https://doi.org/10.1257/jep.14.3.137>
- Palm-Forster LH, Swinton SM, Shupp RS (2017) Farmer preferences for conservation incentives that promote voluntary phosphorus abatement in agricultural watersheds. *J Soil Water Conserv* 72:493–505. <https://doi.org/10.2489/jswc.72.5.493>
- Palm-Forster LH, Suter JF, Messer KD (2019) Experimental evidence on policy approaches that link agricultural subsidies to water quality outcomes. *Am J Agric Econ* 101:109–133. <https://doi.org/10.1093/ajae/aay057>
- Parkhurst GM, Shogren JF (2007) Spatial incentives to coordinate contiguous habitat. *Ecol Econ* 64:344–355. <https://doi.org/10.1016/j.ecolecon.2007.07.009>
- Parkhurst GM, Shogren JF (2008) Smart subsidies for conservation. *Am J Agric Econ* 90:1192–1200. <https://doi.org/10.1111/j.1467-8276.2008.01203.x>
- Parkhurst GM, Shogren JF, Bastian C et al (2002) Agglomeration bonus: an incentive mechanism to reunite fragmented habitat for biodiversity conservation. *Ecol Econ* 41:305–328. [https://doi.org/10.1016/S0921-8009\(02\)00036-8](https://doi.org/10.1016/S0921-8009(02)00036-8)
- Poe GL, Schulze WD, Segerson K et al (2004) Exploring the performance of ambient-based policy instruments when nonpoint source polluters can cooperate. *Am J Agric Econ* 86:1203–1210
- Reeling C, Palm-Forster LH, Melstrom RT (2018) Policy instruments and incentives for coordinated habitat conservation. *Environ Resource Econ*. <https://doi.org/10.1007/s10640-018-0304-2>
- Reddy SMW, Montambault J, Masuda YJ et al (2017) Advancing conservation by understanding and influencing human behavior. *Conserv Lett* 10:248–256. <https://doi.org/10.1111/conl.12252>
- Schultz PW (2014) Strategies for promoting proenvironmental behavior: lots of tools but few instructions. *Eur Psychol* 19:107–117. <https://doi.org/10.1027/1016-9040/a000163>
- Shogren JF, Taylor LO (2008) On behavioral-environmental economics. *Rev Environ Econ Policy* 2:26–44. <https://doi.org/10.1093/reep/rem027>
- Shogren JF, Parkhurst GM, Banerjee P (2010) Two cheers and a qualm for behavioral environmental economics. *Environ Resour Econ* 46:235–247. <https://doi.org/10.1007/s10640-010-9376-3>
- Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* 22:1359–1366. <https://doi.org/10.1177/0956797611417632>
- Smaldino PE, McElreath R (2016) The natural selection of bad science. *R Soc Open Sci* 3:160384. <https://doi.org/10.1098/rsos.160384>
- Spraggon J (2004) Testing ambient pollution instruments with heterogeneous agents. *J Environ Econ Manag* 48:837–856. <https://doi.org/10.1016/j.jeem.2003.11.006>
- Spraggon JM (2013) The impact of information and cost heterogeneity on firm behaviour under an ambient tax/subsidy instrument. *J Environ Manag* 122:137–143. <https://doi.org/10.1016/j.jenvm.2013.02.032>
- Spraggon J, Oxoby RJ (2010) Ambient-based policy instruments: the role of recommendations and presentation. *Agric Resour Econ Rev* 39:262–274
- Stuart D, Benveniste E, Harris LM (2014) Evaluating the use of an environmental assurance program to address pollution from United States cropland. *Land Use Policy* 39:34–43. <https://doi.org/10.1016/j.landusepol.2014.03.009>
- Sturm B, Weimann J (2006) Experiments in environmental economics and some close relatives. *J Econ Surv* 20:419–457. <https://doi.org/10.1111/j.0950-0804.2006.00285.x>
- Suter JF, Vossler CA (2014) Towards an understanding of the performance of ambient tax mechanisms in the field: evidence from upstate new york dairy farmers. *Am J Agric Econ* 96:92–107. <https://doi.org/10.1093/ajae/aat066>
- Suter JF, Vossler CA, Poe GL, Segerson K (2008) Experiments on damage-based ambient taxes for non-point source polluters. *Am J Agric Econ* 90:86–102

- Suter JF, Duke JM, Messer KD, Michael HA (2012) Behavior in a spatially explicit groundwater resource: evidence from the lab. *Am J Agric Econ* 94:1094–1112. <https://doi.org/10.1093/ajae/aas058>
- Suter JF, Collie S, Messer KD et al (2018) Common pool resource management at the extensive and intensive margins: experimental evidence. *Environ Resour Econ*. <https://doi.org/10.1007/s10640-018-0283-3>
- Thaler RH, Sunstein CR (2008) *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press, New Haven
- Thalheimer W, Cook S (2002) *How to calculate effect sizes from published research: a simplified methodology*. Work-Learning Research, Somerville
- U.S. Environmental Protection Agency (2018) National summary of state information water quality assessment and TMDL information. https://ofmpub.epa.gov/waters10/attains_nation_cy.control#total_asses sed_waters. Accessed 27 Aug 2018
- Vossler CA, Poe GL, Schulze WD, Segerson K (2006) Communication and incentive mechanisms based on group performance: an experimental study of nonpoint pollution control. *Econ Inq* 44:599–613. <https://doi.org/10.1093/ei/cbj043>
- Wallander S, Ferraro P, Higgins N (2017) Addressing participant inattention in federal programs: a field experiment with the conservation reserve program. *Am J Agric Econ*. <https://doi.org/10.1093/ajae/aax023>
- Xie J, Cai TT, Maris J, Li H (2011) Optimal false discovery rate control for dependent data. *Stat Interface* 4:417–430
- Yekutieli D (2008) Hierarchical false discovery rate-controlling methodology. *J Am Stat Assoc* 103:309–316. <https://doi.org/10.1198/016214507000001373>
- Zarghamee HS, Messer KD, Fooks JR et al (2017) Nudging charitable giving: three field experiments. *J Behav Exp Econ* 66:137–149. <https://doi.org/10.1016/j.socec.2016.04.008>
- Zhang L, Ortmann A (2013) *Exploring the meaning of significance in experimental economics*. Social Science Research Network, Rochester

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Leah H. Palm-Forster¹  · Paul J. Ferraro²  · Nicholas Janusch³ · Christian A. Vossler⁴ · Kent D. Messer¹

¹ Department of Applied Economics and Statistics, University of Delaware, 531 S. College Ave., Newark, DE 19716, USA

² Bloomberg School of Public Health, Carey Business School and Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA

³ Energy Assessments Division, California Energy Commission, Sacramento, CA, USA

⁴ Department of Economics and Howard H. Baker Jr. Center for Public Policy, University of Tennessee, Knoxville, TN, USA