# The Impact of Experiments on Environmental and Natural Resource Policy

Christian A. Vossler*, Timothy N. Cason†, James J. Murphy‡, Paul J. Ferraro§, Todd L. Cherry‖, George Loewenstein#, Peter Martinsson**, Jason F. Shogren††, Leaf van Boven‡‡, and Daan van Soest§§

## Introduction

Al Roth famously suggested that an important use of experimentation in economics is to "whisper in the ears of princes" (Roth 1986). He was referring to laboratory experiments with human participants designed to test features of real world public policies or to explore the likely consequences of proposed policies. By "wind-tunneling" or "test-bedding" public policies in a rigorously controlled environment, experimental methods can help evaluate policy proposals and test new institutional designs (Smith 1994). Experiments can provide insights into citizen responses to policies and can improve the design of stated preference (SP) studies (an experimental, survey-based approach for valuing public goods) used to inform regulatory analysis (Shogren and Taylor 2008; Carlsson and Johansson-Stenman 2012). In this review, we document the use of experimental findings by government agencies and other public

*Department of Economics, University of Tennessee (corresponding author; email: cvossler@utk.edu); †Department of Economics, Purdue University; ‡Department of Economics, University of Alaska Anchorage; §Carey Business School and Department of Environmental Health and Engineering, Johns Hopkins University; ‖Department of Economics, University of Wyoming; #Department of Social and Decision Sciences, Carnegie Mellon University; **Department of Technology, Management and Economics, Technical University of Denmark; ††Department of Economics, University of Wyoming; ‡‡Department of Psychology and Neuroscience, University of Colorado; §§Department of Economics, Tilburg University

This article was made possible through the input and efforts of a large group of academics, administrators, and public sector employees, who collectively helped us identify and substantiate links between experimental work and policy-relevant outcomes. The authors thank Spencer Banzhaf and two anonymous referees for feedback that significantly improved the article. The article benefited from discussion and suggestions provided by participants in the 2022 and 2023 Teton Group Workshops held in Jackson, Wyoming. CAV, TNC, JJM: conceptualization, investigation, writing—original draft; PJF, TLC, GL, PM, JFS, LvB, DvS: conceptualization, investigation, writing—review and editing.

Electronically published June 16, 2026

*Review of Environmental Economics and Policy*, volume 20, number 2, summer 2026.
 Published by The University of Chicago Press for AERE. https://doi.org/10.1086/741736

organizations to inform the design of new environmental and natural resource policies or to amend the structure of existing policies.[1]

The policy impacts of experiments are rarely traceable by examining the original papers and are difficult to locate via targeted literature searches. As such, we gathered much of the evidence we present from personal communications with researchers at academic institutions, administrators in government agencies, and crowdsourcing efforts through organizational listservs. Other sources include policy briefs and websites that provide easily accessible and digestible information on research projects and their policy impacts. Although some experimental results were employed in making specific policy design choices, in far more instances, experiments provided only broad guidance for improving policy design or supporting policy action.

There have been targeted efforts to connect academic research, including research involving SP methods, with its use in benefit–cost analysis (BCA; e.g., Banzhaf 2010; Griffiths et al. 2012; Atkinson et al. 2018). Normann and Ricciuti (2009) assess the contributions of laboratory experiments to policy-making and provide some examples of how experiments have affected policies (including some related to the environment). Relative to these prior efforts, we consider the full spectrum of experimental methodologies and a broader range of environmental and natural resource applications. Nevertheless, to provide sufficient depth on the examples we highlight, we do not include all the evidence we uncovered. Moreover, in some cases, although our sources are convinced that results from experiments had influenced policy decisions, they could not document the links, typically because they had lost contact with decision makers or failed to retain documentation (e.g., emails, reports, etc.). This review should therefore not be considered comprehensive.

By documenting many cases in which experimental research has informed policy, we aim to demonstrate its practical relevance and persuade policy makers—particularly those who may be skeptical of the potential of experimental research—of its value. Relevant and transparent examples can convince policy makers, funders, public organizations, students, and researchers of the value of experimentation and can provide an understanding of how to use its insights in policy (e.g., see Duflo 2018; Shogren et al. 2021). We conclude by making recommendations to experimental researchers about how to increase the policy impact of their work.

## Experimental Methods and Sphere of Influence

Formal experimentation is a central component of science; before proceeding, it is important to be clear what we mean by an "experiment." We define an experiment broadly to mean any

[1]Outside the environmental context, high-profile examples of experiments used by US government agencies to inform policy decisions include the Treasury auctions in the 1970s (Smith 1994), payload allocation on NASA's Cassini mission to Saturn (Ledyard, Porter, and Wessen 2000), the Securities and Exchange Commission's analysis of short sale price restrictions (US SEC, Office of Economic Analysis 2007), and the design of the Federal Communication Commission's spectrum auctions (Plott 1997). Evidence-based experimental approaches are required for the adoption of new medical procedures and medications and are frequently incorporated into the evaluation of social and education programs, such as the Oregon Medicaid program (Finkelstein et al. 2012) and the Tennessee Student/Teacher Achievement Ratio experiment (Krueger 1999).

scientifically controlled (i.e., a nonnaturally occurring) procedure intended to examine the effect of a mechanism, institution, or intervention on human behavior, or to measure preferences for potential policies. This definition excludes studies typically referred to as "natural experiments" or "quasi-experiments." Unlike these studies, most studies we document employ random treatment assignment with control groups, which provides a gold standard for causal identification.[2]

Our definition of an experiment encompasses a wide variety of methodologies, including highly stylized experiments conducted in university laboratories using student subjects, lab-in-the-field experiments employing industry professionals or consumers as participants, and field experiments implemented in natural settings. We also include SP surveys that use experimental methods to measure preferences and predict behavior. The most appropriate experimental methodology and subject pool depend on the research questions asked and types of policy guidance desired (Henrich, Heine, and Norenzayan 2010; Cason and Wu 2019).

One of the most useful roles that experiments can play is in the policy design stage. Economic theory can guide policy design, but empirical evidence provides greater confidence that a policy will accomplish the policy makers' objectives.[3] Experiments are especially useful when considering a novel policy innovation where, typically, no naturally occurring data exist that indicate whether the policy would work as intended. Such "testbed" experiments are particularly influential when policy makers engage with experimenters at an early stage, because the testbed experiments can help refine specific policy implementation details.

Experiments can also be useful following policy implementation, by helping assess whether a specific policy feature has led to intended or unintended results and by suggesting improvements to the policy design.[4] Making changes after policy implementation can, however, be challenging. It may be legally difficult to make changes, especially when aspects of the initial policy design were guided by enabling legislation, and because policy changes often require an extensive notification and comment period.

Although experiments can provide useful input to decision-making, we acknowledge that complementary evidence is generally needed to justify a specific policy.[5] An experiment can provide direct evidence of causal links and provide estimates of the impact of a policy. However, replication of the results of the experiment, as well as complementary analyses, such as an assessment of benefits and costs, may be needed prior to implementation. Even when a causal link between a policy and a desired outcome is clear, this alone does not justify the policy intervention. Experiments typically, and intentionally, overlook a range of other considerations that are likely to play an important role in policy adoption, such as political,

[2]For a critical discussion on randomized control trials (RCTs), see Deaton (2020).

[3]Del Carpio, Alpizar, and Ferraro (2021), in the context of a monitoring program in rural Costa Rica aimed at reducing community water use, provide an example of how researchers and policy makers can collaborate during the experimental design phase to ensure a common understanding of what is being studied, how outcomes will be measured, and how hypotheses will be tested.

[4]For example, policy makers may rely on engineers' estimates of resource savings. However, if people cannot use the technology correctly, then the actual gain may be much smaller (Alpizar, Del Carpio, and Ferraro 2024).

[5]We thank Dale Whittington for emphasizing this point to us.

ethical, and fiscal concerns (e.g., the political fallout reported in Browne et al. [2025], discussed later). Moreover, some interventions that are successful in small trials may not work when implemented at a larger scale (see Al-Ubaydli et al. [2021] and List [2024] for a discussion of scaling problems and ways to mitigate them). These and other considerations are more likely to be captured in quasi-experiments and nonexperimental research (e.g., historical analysis and event studies), which make complementary contributions to policy design and implementation (Connolly, Loewenstein, and Chater 2025). In the next sections, we describe experiments that informed policy decisions.

## Property Rights Trading

Laboratory experiments have provided testbeds to improve the performance of property rights trading programs, including cap-and-trade systems for controlling air emissions, individual transferable quota schemes to regulate fishing, and markets for trading water allocations. Such programs can achieve policy objectives at a lower cost than traditional command-and-control policies—for example, regulations that require the use of specific pollution control technologies. In the case of emissions trading programs, regulators set a cap on total emissions, and firms are allocated (or purchase) permits (also called allowances) to emit a specified quantity of emissions. Firms can trade these permits, creating a financial incentive to reduce emissions where it is cheapest to do so. To implement these and other trading programs, policy makers must develop new market institutions and rules for which historical observational data provide little guidance. However, these new institutions and rules can be varied in experiments to evaluate their empirical consequences.

As an early example, a testbed experiment helped to design important features of the Regional Clean Air Incentives Market (RECLAIM) trading program in Southern California. The South Coast Air Quality Management District (AQMD) used emissions trading to reduce nitrogen oxides ($NO_x$) and sulfur dioxide ($SO_2$) emissions from stationary sources (such as power plants) by annually phasing down allocated permits. In trading programs, permits are associated with a specific compliance period, which has a defined start and end date. When too many or too few permits are available late in a compliance period, the permit prices in the market can be volatile. The most common way to limit permit price volatility is by allowing traders to bank permits for later use, but the RECLAIM program did not allow the banking of unused permits for use beyond their compliance year. To provide an alternative way of smoothing out price variability without banking, Carlson et al. (1993) developed a set of experiments to test a new idea. In a pair of experiments, the researchers showed that price volatility arising from a uniform permit expiration date could be nearly eliminated by staggering expiration dates. Doing so helped avoid the "squeeze" of excess supply or demand at the uniform expiration date. The AQMD adopted the staggered expiration date design, which continues to this day.

One way to allocate emissions permits is through auctions. Early testbed experiments were motivated by a new auction institution designed by EPA when implementing Title IV of the 1990 Clean Air Act Amendments. The goal of this comprehensive, nationwide emissions trading program was to reduce $SO_2$ emissions by about one-half. The new auction had unusual incentive properties, in that individual sellers received the bid prices of "matched" buyers, and higher bid prices were given to sellers who offered lower prices.[6] All individual bids and offers thus directly affected transaction prices. Cason (1993) showed theoretically how this creates incentives for sellers to submit offers below their cost of emission control. This can generate downwardly biased price signals and inefficient trading, unrealized mutually beneficial trades, and seller losses.

Although the unusual bid-to-offer matching creates undesirable incentives in theory, it is an empirical question whether human bidders' behavior changes because of these incentives. A laboratory experiment by Cason (1995) found that seller offer behavior was broadly consistent with the theoretical incentives of the new EPA auction. In a subsequent experiment, Cason and Plott (1996) compared the behavior and performance of the EPA auction rules with the more common uniform-price auction rules, which are used daily by specialists on the New York Stock Exchange to set opening prices. Buyers in the EPA auction bid below their true permit valuations to obtain greater surplus when buying, as is standard in pay-as-bid auctions. Sellers in this auction recognized that offering permits at lower prices allowed them to match up with higher bids and receive higher prices. After several decision rounds, most sellers submitted offers well below their emission control costs. The level of value and cost misrepresentation by buyers and sellers was much greater in the EPA auction than in the more standard uniform-price auction, where only marginal bids and offers affect prices. Consequently, prices were less accurate in reflecting the underlying emission control costs.

From the evidence provided by these experiments, as well as low early participation in the EPA auction, the US General Accounting Office recommended switching to the more common uniform-price rules (US GAO 1994). EPA attorneys determined that the statutory language of the act permitted this change, and the agency proposed making that switch because the new price rule "may have the benefit of producing a less confusing price signal" (US EPA 1996). EPA never changed the auction rules, however, because they largely abandoned the annual permit auction for interfirm trading. Moreover, by 2010, the $SO_2$ permits lost all value because of other regulations that limited emissions. Nevertheless, the experimental testbeds did encourage EPA to rethink its auction pricing rules (Schmalensee and Stavins 2013, 2017).

In other emissions trading programs, regulators engaged with experimentalists to assist in the market design process. For example, Virginia's Department of Environmental Quality (DEQ) commissioned a series of experiments (Porter et al. 2009) to help design their auction for $NO_x$ allowances. The experiments compared the revenue and efficiency of three auction mechanisms, known as a sequential English clock auction, a combinatorial sealed-bid auction, and a combinatorial English clock auction.[7] The combinatorial auction formats can account for the potential substitutability between allowances of different vintages, allowing bidders to submit "any/or" bids specifying acceptable prices and quantities for different combinations of vintages.[8] The DEQ sought an auction design that would maximize revenue and

[6]Clean Air Act Amendments of 1990 (Public Law 101-549), Sec. 416(d)(2).

[7]The English clock auctions use a clock to quote successive prices, and each bidder indicates the quantity they demand at the standing price. The auction ends when the quantity demanded equals the supply of allowances. In the sealed-bid auction, allowances are allocated to the successful bidders on a pay-as-bid basis.

[8]The vintage of the permit defines the year when it can first be used. A 2004 vintage allowance can be used in 2004 and in the future (subject to restrictions), but a 2005 vintage allowance cannot be used in 2004.

ensure that permits went to firms that value them the most. The results indicated that the sequential English clock auction generated the highest revenues, while also being simpler than the combinatorial alternatives, and the Virginia DEQ chose to implement this auction format. The timing of the experiments and Virginia's auction adoption likely influenced other signatory states of the Regional Greenhouse Gas Initiative (RGGI) to shift toward allowances auctioning. The experiments served as a bridge to policy implementation for the well-developed theoretical literature on auctions.

In another example of regulators engaging with experimentalists, the New York State Energy Research and Development Authority funded an extensive series of auction design experiments for selling carbon dioxide ($CO_2$) emission allowances in RGGI (Holt et al. 2007). The experiments compared different auction formats across a set of performance measures, including whether auction prices reflected their market value following an unanticipated shift in demand for allowances. The study recommended a uniform-price sealed-bid auction, which was adopted for RGGI, as well as subsequently by the EU Emissions Trading System (EU ETS) and the California greenhouse gas market (Assembly Bill 32).[9] This recommendation was supported further by experiments in which bidders could tacitly or explicitly collude in uniform and discriminatory (i.e., prices match the offers of sellers) sealed-bid and clock auctions (Burtraw et al. 2009).

More recently, two sets of experiments made recommendations for changes to permit auction rules, which were subsequently adopted in emissions trading programs. The first set of experiments tackled a common problem in greenhouse gas emissions trading programs, in which permits are oversupplied in the early years, leading to extremely low prices (Narassimhan et al. 2018). To address this problem, Burtraw et al. (2022) demonstrated the effectiveness of using a price-responsive supply schedule that automatically adjusts the supply of allowances to new information about abatement costs. The experiment provides evidence that this mechanism increases allowance prices and reduces volatility in prices and revenues. This supply adjustment process was adopted recently by RGGI and is being considered by the California market.[10] Holt and Shobe (2016) report an experiment that evaluated the effectiveness of a quantity collar, which adjusts the auction quantity based on the stock of banked permits and is very similar to the way in which the EU ETS was adjusted in 2019 when introducing the Market Stability Reserve. Interestingly, Holt and Shobe (2016) compared the effectiveness of such a quantity collar with a policy with a system of minimum and maximum permit prices—a price collar. Comparing the quantity and the price collar, the latter reduces price variability and increases efficiency, but the experiment results arrived too late in the policy-making process for the EU ETS to take these results into account.

Betz et al. (2010, 2017) report experiments that tested different auction designs for $CO_2$ permit allocation in Australia, comparing sealed-bid and alternative clock formats, as well as simultaneous and sequential bidding. The clock format performed best in simultaneous auctions, but switching to a sequential procedure gave an efficiency advantage to

[9]Personal communication with Charles Holt (May 3, 2023).
[10]Personal communication with Charles Holt (May 3, 2023).

sealed-bid rules. The draft legislation that created the $CO_2$ trading scheme followed closely the recommendations based on the experimental results (Australian Government 2013). However, after the 2013 election, a new coalition government abolished the carbon pricing mechanism.

The Australian government also used experiments to help design a credit trading trial to address dryland salinity, comparing rules based on uniform price (i.e., same price to all sellers) and discriminatory price (i.e., prices match the offers of sellers) and different information conditions during bidding (Connor et al. 2008). Results led to the adoption of closed-book, uniform-price rules for field implementation, and BCA estimated net positive benefits for the program overall.

## Conservation Auctions

Experiments have also informed the design of conservation auctions, which allocate funds to pay landowners to implement conservation practices. One important example is the Georgia Irrigation Reduction Auction. In 2000, Georgia passed a law that required the state's Environmental Protection Division (EPD) to use an unspecified "auction-like process" to pay some farmers to suspend irrigation in drought years. In the auction, farmers would compete for payments from the EPD to voluntarily cease irrigating their land from a well for the remainder of the calendar year.

To testbed alternative auction designs, the EPD contracted with experimentalists (Cummings, Holt, and Laury 2004). Given that farmers can communicate before and during an actual auction, the researchers sought a "collusion-proof" design and allowed face-to-face communication during the experiment. Moreover, EPD officials wanted the auctions to allow farmers to revise offers. Among the design issues investigated were the tie-breaking rule (in the event a tie occurs at the highest acceptable price), uniform- versus discriminatory-pricing rules, and the approach for announcing accepted offers each round (by permit ID or by announcing the highest accepted price). To help EPD officials better understand and trust the results of the experiments, the experimentalists invited the EPD director and staff to attend a laboratory session and to review the study's findings. The EPD implemented the experimenters' recommended procedures: a discriminatory-price auction with opportunities for sellers to revise their offers, no announcement of the maximum accepted price, and a random tie-breaking rule (Cummings, Holt, and Laury 2004). To pursue noneconomic objectives, however, the EPD director did not apply a consistent rule for provisionally accepting offers across revision rounds.

In Victoria, Australia, the Department of Natural Resources and Environment sought bids for projects to improve native vegetation management on private land. To help design these BushTender auctions, the department commissioned a series of laboratory experiments (Cason, Gangadharan, and Duke 2003). In conservation auctions that differentiate project offers by their environmental benefits, an important design decision is whether the regulator should reveal the benefit information to sellers. Providing this information could improve auction performance if it encourages landowners to submit offers closer to their true costs. However, landowners who learn that their projects are highly valued may respond by submitting higher offers. In the experiment, sellers receiving benefit information misrepresented their costs more for high-benefit projects. Consequently, the overall environmental benefits acquired in the laboratory conservation auctions were lower (and seller profits were higher). The experimental results thus led to a recommendation to withhold benefit information from sellers, and the government followed this advice when it implemented the policy in Victoria (Stoneham et al. 2003). Analysis of the auction data led to estimates of substantial cost savings relative to fixed-rate payments, which were the prevailing mechanism for subsidizing conservation activities at the time (Schilizzi and Latacz-Lohmann 2007).

Balmford et al. (2023) used a combination of prior auction theory and lab experiments to persuade an industry partner to change a conservation auction from a discriminatory-price rule (pay-as-bid) to a uniform-pricing rule. The specific auction allocates contracts to farmers located in the Poole Harbour catchment within the United Kingdom to plant cover crops to reduce downstream nitrogen runoff. Their analysis focused on a comparison of auction behavior and outcomes following this switch with earlier results using pay-as-bid rules, as well as a "control" catchment that maintained these older pricing rules. The positive results from this field experiment (lower bids, lower payments to farmers, and less strategic behavior) led to continued use of uniform-pricing rules in the tested catchment, as well as the adoption of these rules as the pricing default when these conservation schemes expanded geographically in the United Kingdom.

Auctions can also encourage bidders to engage in spatial coordination, which may increase the ecological benefits of conservation efforts (for a review, see Hanley 2026). Rolfe, McCosker, and Windle (2008) describe pilot tests conducted with landholders. The goal was to create "corridors" protecting native vegetation, which required coordination among adjacent parcels with different owners. Insights from the experiment were used in a full-scale rollout in Queensland, Australia, leading to an auction design with multiple bidding rounds to help landholders learn about the process and its risks and to provide feedback on the locations of other bid proposals (Windle et al. 2009).

## Environmental Programs Directed at Individuals and Households

Experiments have been important to the development of interventions that encourage individuals and households to limit their use of natural resources and participate in environment-related programs. Some of these studies develop and evaluate the impact of information "nudges" (Thaler and Sunstein 2008). For example, Schultz et al. (2007) studied the effects of normative information on residential energy use. They find that providing information on average electricity usage decreased use among above-average energy consumers and increased use among below-average consumers. For the latter group, the undesirable effect of increased electricity use was offset when this messaging was combined with a happy face emoticon. This article influenced companies that work with public utilities, such as Opower (see Schultz et al. 2018) and VertexOne (previously WaterSmart), to use social norms to encourage conservation behavior. A report by WaterSmart (2014) documents that they scaled up the Schultz et al. (2007) methodology and sent "Home Water Reports" that reduced water consumption by up to 5 percent; Opower interventions cut energy usage by 2–5 percent (Schultz et al. 2018).

The European Commission Directive 2009/72/EC requires EU member countries to evaluate the rollout of smart electricity meters (which allow consumers to digitally monitor energy use in real time) and to equip at least 80 percent of customers by 2020, subject to a favorable BCA. To meet these objectives, Ireland's Commission for Energy Regulation, in coordination with various industry partners, conducted separate randomized controlled trials (RCTs) for electricity and natural gas to test the effects of detailed energy use statements, billing frequency, smart meters, and pricing schemes that vary depending on the time of day (CER 2011a, 2011b). For both energy sources, the field trials showed that using a combination of a detailed use statement, bimonthly billing, and smart meters resulted in significant reductions. Encouraged by these results, Ireland's National Smart Metering Programme began in 2019.

In principle, improvements in home energy efficiency offer the potential to reduce homeowners' energy bills while also lowering greenhouse gas emissions. However, most policies designed to improve home energy efficiency have not met expectations (Levinson 2016). As part of the efforts of the Energy Policy Institute at the University of Chicago (EPIC) to provide evidence-based recommendations to the US government (EPIC 2021a), Burlig (2021) offered recommendations on how to improve the performance of home energy efficiency programs, based in part on experimental evidence by Allcott and Greenstone (2017) and Fowlie, Greenstone, and Wolfram (2018). These recommendations include requiring federal programs to incorporate real world evaluation to identify which retrofits have the highest return, as well as to identify implementation failures. As documented in EPIC (2021b), the Infrastructure Investment and Jobs Act established a grant program that funds commercial and residential energy audits, as well as upgrades and retrofits. Within a year after the upgrades are made, a final audit is required to assess the total energy savings, reflecting Burlig's recommendation.

The Behavioural Insights Team (BIT), originally created by the UK government but now operating as an independent consultancy, applies a behavioral science approach to address public policy problems.[11] One environment-related issue that BIT addressed with experiments was the sustainable consumption of electronic and household electric appliances. BIT used laboratory experiments with consumers to help the French government develop a repairability index, which informs consumers about the repairability of electronics and electric products, and then evaluated the most promising options with an RCT (BIT 2024). The repairability index came into force in 2021. A post-implementation evaluation suggested an increase in the share of more repairable products sold, and producers also responded by building newer models that were more repairable (BIT 2024).

In a second project, the BIT partnered with Chicago's Department of Water Management (DWM) to design and test, with an RCT, methods for increasing return rates for people who had requested a drinking water test kit to detect lead (BIT 2022). One intervention redesigned the test kit to make it easier for people to complete the test, and a second sent text message reminders to return the kit. The new kit increased return rates by more than 20 percentage

[11]The existence of more than 300 Behavioral Insights units across the globe (https://oecd-opsi.org/bi-units/) exemplifies the growing interest in using evidence from experiments and other methods to inform policy.

points, with an additional 3.8 percentage point increase among those who received a reminder. The results motivated the DWM to adopt the new kits and to send text reminders (BIT 2022).

## Regulatory Enforcement and Compliance

In the context of regulatory enforcement to promote compliance, several experiments have led to the adoption or modification of rules. These changes affected both consumers and producers. On the consumer side, a recent example concerns the enforcement of restrictions on residential outdoor water use during drought conditions in California. Although smart water meters reduce ongoing monitoring costs to near zero and have been available in some areas for over a decade, many communities continue to rely on costly in-person monitoring. In collaboration with the city of Fresno, researchers implemented a field experiment to examine the effects of smart meters that could perfectly detect violations of outdoor water use restrictions (Browne et al. 2025). By randomizing whether households were subject to automatic enforcement, as well as the levels of the fine for violations, they found that automated enforcement reduced violations by 17 percent and decreased water consumption by about 3 percent. However, automated enforcement also increased customer complaints by 1,100 percent. Many of these complaints arose because of a lack of understanding about the policy's motivation, ultimately leading to the program's cancellation. This suggests that better communication could have mitigated these concerns and highlights that political considerations can limit technological solutions to enforcement challenges.

In another consumer context, a field experiment shed light on the effectiveness of using a threat of punishment to encourage households to sort their waste (Vollaard and van Soest 2024). The experiment randomized the timing of a warning letter, followed by a 1-month period of highly visible and intensive law enforcement, across 65 waste collection routes in Tilburg, Netherlands. These interventions led to a persistent drop in residual waste of 10–15 percent. The experiment identified the favorable financial performance of this enforcement program, which led to its implementation and expansion to additional areas (van Oirschot 2015).

On the industry side, Duflo et al. (2013) report on a 2-year experiment conducted in Gujarat, India, that varied the incentives of auditors for environmental performance. Gujarat is a fast-growing industrial state that suffers from some of the highest levels of industrial pollution in the country. In 1996, the authorities introduced a third-party environmental audit system, with auditors hired by the plant. The system was functioning poorly, with regulators viewing the audits as useless because of inaccuracies. Researchers collaborated with regulators to modify the audit and inspection system to improve its accuracy. In the study, half of the auditors were hired and paid by the plants, which was the status quo, and the other half were paid fixed fees and received incentive payments for accurate reports. Results indicated that the status quo system was clearly corrupt, as auditors systematically reported emissions marginally below the standard, even though actual emissions were generally higher. In the experimental treatment, auditors reported more truthfully, and this significantly reduced the false compliance rate; consequently, plants receiving the experimental treatment reduced their pollution emissions. The state used these results to improve monitoring and enforcement procedures. Air and water pollution at plants receiving the new audit format were significantly lower within 6 months, reducing air pollution by 28 percent.[12]

Benami et al. (2023) report on an experiment that studied how text message reminders can improve on-time submission of discharge reports for compliance with the US Clean Water Act. Collaborating with the US EPA in this information nudge experiment, the researchers customized messages ahead of reporting deadlines to enhance the firms' perceived probability of violation detection. Although treatment effects were heterogeneous, the pre-deadline reminders increased overall on-time submissions by almost 7 percentage points. The administrators of the EPA's office responsible for receiving electronic discharge monitoring reports indicated that, based on this research, the system's functionality was adjusted to issue pre-deadline reminders for report submission to enhance compliance.[13]

## Nonmarket Demand for Environmental Goods and Services

A significant contribution of environmental economics to policy has been the development of tools for assigning a monetary value to the benefits of environmental and other nonmarket goods. SP experiments are the leading nonmarket valuation approach. They can place a value on policy outcomes that go beyond the scope of observational data. They can also quantify important sources of both use value (such as value from recreational fishing) and nonuse value (such as value placed on the mere existence of specific ecosystems) that are associated with changes in environmental goods and services. SP surveys mainly enter the policy arena through government BCA, which policy makers often use to justify environmental regulations. BCA is generally an input to decision-making rather than a stand-alone decision tool. Nevertheless, its global use and the frequent inclusion of SP estimates in BCA indicate vast potential for influence.[14]

SP surveys have received considerable skepticism because they rely on potential behavior—what people say they are willing to do or pay—rather than actual behavior. These concerns have led to a large body of research.[15] Important catalysts for methodological developments included early research funded by EPA, Ronald Reagan's 1981 executive order requiring BCA for major regulations, and the National Oceanic and Atmospheric Administration (NOAA) Panel on Contingent Valuation's qualified endorsement of SP surveys (Arrow et al. 1993).[16] Laboratory and field experiments have been key to these developments. Such experiments have been used, for example, to compare alternative procedures for eliciting the value that survey participants place on environmental costs and benefits, and for testing the efficacy

[12]https://urbanlabs.uchicago.edu/projects/gujarat-auditing-the-air.

[13]Personal communication with Elinor Benami (September 7, 2023).

[14]A handful of studies summarize the role of nonmarket valuation research, including SP, in various environmental policy contexts spanning the United States (Loomis 2000; Banzhaf 2010; Griffiths et al. 2012; Petrolia et al. 2021), the United Kingdom (Atkinson et al. 2018), and Australia (Baker and Ruting 2014). Some of the research that we elected to highlight is referenced in these related studies.

[15]Most SP valuation scenarios are hypothetical, in the sense that decision makers are generally not considering the exact policies described, nor have they committed to using a decision rule based on survey results. Nevertheless, most participants view their responses to high-quality SP surveys as consequential (e.g., Herriges et al. 2010), and survey results can influence policy decisions, as documented in this article.

[16]An important output from early EPA funding is an assessment of the methodology (Cummings, Brookshire, and Schulze 1986).

of information scripts that encourage participants to provide truthful responses. Experiments have further been important for external validity testing—that is, whether statements of what people say they would do reflect what they actually would do. In particular, experiments provide the opportunity to construct markets for the provision of public goods and vary the financial consequences of the choices presented to participants. These methodological experiments have been critical to increasing the acceptability of SP methods among academics and policy makers.[17]

We first highlight four studies associated with water resource management decisions. A survey by Jones & Stokes Associates (1994), as discussed by Loomis (2000), helped inform the decision to increase water flows into Mono Lake to protect migratory birds by reducing the water rights of Los Angeles. The SP study elicited valuations for different lake elevations and concluded that an elevation of 6,390 feet yielded the highest benefits. The California State Water Resources Control Board (1994) issued a decision that the lake must rise to 6,392 feet, which is close to the research recommendation. In a second example, Bateman et al. (1995) estimated the value of protecting the Norfolk Broads wetland area in England from flooding from the North Sea. A defining feature of the study is that it followed the best practice guidelines proposed by the NOAA panel. The postscript to Bateman, Langford, and Rasbash (2001, 536) states, "Given our findings, the National Rivers Authority chose to use the . . . [willingness to pay] WTP measure in their attempt to argue the case for funding from the UK Treasury. . . . As a consequence the principle of funding flood defences in Broadland has now been accepted and the construction of new defences has recently commenced."

As a third example, Welsh et al. (1995) used a national survey to estimate the total economic value of changes in the operating criteria for Glen Canyon Dam on the Colorado River. Concerns over the impacts on downstream resources led to a decision by the Secretary of the Interior for Reclamation to prepare an environmental impact statement to reevaluate dam operations. Of the nine operational alternatives considered, the Modified Low Fluctuating Flow Alternative was selected because it provided the most benefits. The Record of Decision (US DOI 1996, 13) highlights that, although the operational changes would result in a decrease in hydroelectric power generation, "a recently completed non-use value study . . . indicates that the American people are willing to pay much more than this loss to maintain a healthy ecosystem in the Grand Canyon." Last, a study by Adamowicz, Louviere, and Williams (1994) estimated the recreation impacts of alternative flow scenarios for the Highwood and Little Bow Rivers in Alberta (Canada). This study provides a framework for combining SPs for potential recreation opportunities with data on actual recreation visits; it highlights that SP experiments, through randomization techniques, can make it easier to identify the effects that different factors, such as travel costs, have on recreation demand. Benefits estimates from the study were part of the BCA that supported the Highwood/Little Bow water resource development project. The executive summary of the review panel (NRCB/CEAA 1998) recommended to the federal government that, in part based on expected "social" effects (i.e., nonmarket values), the proposed project is in the public interest.

[17]Johnston et al. (2017) document many examples of SP experiments related to methodology and validity. As evidence of increased acceptability, the revised US federal guidelines for regulatory analysis (OMB 2023) effectively elevate the standing of SP studies and summarize the Johnston et al. (2017) best practice guidelines.

It is evident from these case studies that it is important for SP studies to survey the affected population, describe potential mechanisms for change, and quantify the environmental and other social impacts of a policy design relative to an appropriate counterfactual scenario. Disconnects between a policy application and an SP study provide fodder for rejecting or downplaying benefit estimates (Baker and Ruting 2014). As emphasized by a series of examples described in Banzhaf (2010), which examined the role that nonmarket valuation research has played in the design of land conservation policies, SP was more likely to influence policy when researchers made efforts to directly communicate findings to stakeholders, present quantitative information in ways that help policy makers gauge political support (e.g., by providing information on the distribution of values in the public rather than average or total willingness to pay), and offer qualitative information such as relative policy importance and people's motivations for being for or against a policy action.

Whereas the above examples are tied to relatively small geographic areas and targeted populations, national policies require more substantive evidence. An early example is the EPA-funded study by Carson and Mitchell (1993). As documented by Griffiths et al. (2012), this study served as a primary source for estimating benefits associated with several rules related to surface water quality. The study was well suited for policy analysis because it used a representative sample of respondents and implemented nationwide valuation scenarios as treatments that were linked to the "designated use" standards of the Clean Water Act. In the United Kingdom, the National Water Environment Benefit Survey (NERA-Accent 2007; Metcalfe et al. 2012) is another example of a "workhorse" national study used in various analyses involving program appraisals and assessments. Estimated values can be adapted to fit different geographic and ecological scales and are based on SP valuation scenarios congruent with the European Union Water Framework Directive.

Atkinson et al. (2018) evaluated the extent to which environmental valuation has influenced UK policy decisions. In the case of water policy, the authors document extensive use of SP methods by the environmental agency and private water companies. In some cases, the BCA appears to have been undertaken merely to fulfill policy mandates, whereas in others, the analysis was an input to help determine the preferred policy alternative. As a more direct link between BCA and policy outcomes, under the Water Framework Directive (EC 2000/60/EC), areas with benefit–cost ratios of less than 0.5 (where costs were twice as much as benefits) triggered the possibility of exemption from minimum water quality standards; exemptions have been granted, which resulted in substantial cost savings. In the case of "natural capital" policies—which account for the value of nonmarket natural resources—BCAs have further influenced government budget allocation decisions.

Because of resource constraints, it is often impractical to conduct a case-specific SP study to inform a regulatory analysis. The trend has been to use benefit transfer, which relies on using results from existing nonmarket valuation studies to estimate values for new policy scenarios. Benefit transfer exercises are done primarily through integrated assessment models (e.g., BenMAP, HAWQS, NEVO) and meta-analyses of multiple studies. For example, when determining the benefits associated with decreased mortality rates, many government agencies use results from multiple studies to form a central estimate of the value of a statistical life (VSL). These VSL estimates represent a large share of the benefits of many environmental regulations, including those related to the health benefits of air quality and drinking water standards. On one hand, benefit transfer decreases the need for a single study within the context of a specific policy analysis. On the other hand, a single study often is used to inform myriad analyses. For example, as documented by Petrolia et al. (2021), 14 SP studies related to mortality or morbidity risks were used in 29 or more of the 49 regulatory impact analyses conducted by EPA between fiscal years 2008 and 2019.

SP experiments also have influenced public decisions outside the benefit–cost context. The New England Fishery Management Council and the NOAA regional administrator rely solely on the outputs from a bioeconomic model, a version of which is described in Lee, Steinback, and Wallmo (2017), to set recreational fishing regulations for Gulf of Maine cod and haddock. The bioeconomic model integrates SP-based estimates of recreation demand with an age-structured model of the dynamics of fish stock. In 2023, the same modeling approach was used to set fishing regulations for summer flounder, black sea bass, and scup in the northeastern states. There is an effort at NOAA Fisheries to replicate this modeling approach across the country.[18]

Aadland et al. (2012) conducted a nationwide study to estimate the willingness to pay for a new pass for recreational access to US public lands. Using a sample of people who had previously purchased a National Parks Pass, they estimated that the average household would pay more than $80 to replace the pass they previously purchased, and that expected revenue losses were estimated to be modest at prices of $75 or above. Based on this evidence, the federal agencies set the price of the new pass at $80. The authors report that the projected sales from their calibrated model (at a price of $80) were within 12 percent of the actual sales for the first year.

## The Future Experimental-Environmental Policy Nexus

Although our academic, industry, and government sources reported many experiment-policy links, they were more likely to report promising experimental results that were *ignored* in subsequent policy decisions. For example, the US Department of Agriculture (USDA) staff experimentally tested the effects of behavioral nudges, in particular tailored reminder letters, on participation in a Conservation Reserve Program auction known as the General Sign-Up, through which landowners can submit offers to enroll land into conservation practices (Wallander, Ferraro, and Higgins 2017). The lead researchers calculated that the intervention had a benefit–cost ratio of between 20:1 and 90:1. Yet, USDA staff did not use the intervention in subsequent auctions.[19] More broadly, our sources reported that experimental evidence was not routinely used to inform the design of environmental policies and that experimentation was rarely embedded into environmental programs for the purpose of learning and evaluation (observations also made in Ferraro et al. 2023).

The challenge of encouraging practitioners to generate and use experimental evidence is not unique to environmental and natural resource policies. The field of medicine, for example, has a long history of generating experimental evidence. Yet recent concerns about the lack of use of that evidence among practitioners, and the relevance of that evidence under real world conditions have given rise to the field of "implementation science." Using experimental

[18]Personal communication with Scott Steinback (July 27, 2023).
[19]Personal communication with Paul Ferraro (June 18, 2025).

designs, implementation science researchers study the effects of medical interventions under real world conditions (Brown et al. 2017) and test ways of promoting the use of evidence among practitioners (Foy et al. 2023). In areas outside of medicine, researchers are studying how practitioners use evidence (e.g., Vivalt and Coville 2023; DellaVigna, Kim, and Linos 2024) and employing experimental methods to test new ways of encouraging its use (e.g., Hjort et al. 2021; Scott et al. 2023).

We believe that an essential but untapped use of experimentation is to test new ideas to encourage environmental and natural resource policy makers to employ experiments and experimental evidence. For example, academics have written survey articles that are intended to communicate experimental results in accessible ways, for instance, by providing "lessons for policy makers" (e.g., Noussair and van Soest 2014; List and Price 2016; Stranlund 2017). Other forms of accessible communication include nontechnical policy briefs and internet media, such as podcasts and blogs, that can be easily found through search engines. To test whether such forms of communication increase the use of evidence, experimentalists could vary decision makers' exposure to different forms of communication about evidence under both laboratory and field conditions.

Our own experiences, and those of our industry, government, and academic sources, also highlight the importance of efforts to build and maintain relationships between experimentalists and policy implementers. The windows for influencing the design or redesign of policies and programs arise suddenly, and they close just as quickly. Experimentalists are likely to miss those windows if they do not have well-established relationships with implementers and if they are not able to act quickly. Well-established relationships can also help policy implementers develop an appreciation for the power of experimentation and thus engender a greater willingness to use experimental evidence in shaping policies. Such relationships also make it more likely that experimentalists will consider policy-relevant issues when designing experiments, such as whether or how a new program will work when scaled up (List 2024). Equally importantly, close relationships with implementers can help researchers better understand the research needs of various programs and thus identify and disseminate the most relevant research results (e.g., Moore et al. 2023). For example, in the United States, the Foundations for Evidence-Based Policymaking Act of 2018 encourages federal agencies to create annual learning agendas that identify questions for which answers would improve program impacts (described in more detail in Ferraro et al. 2023). The more that experimentalists are involved in crafting these learning agendas with program staff, the more likely it is that experimentation will be used in fulfilling these agendas.

There are no silver bullets for effectively encouraging policy makers to rely more on experimentation and experimental results in the design and implementation of environmental and natural resource policies. However, we believe that our documentation of previous productive links between experiments and policies warrants continued efforts to find ways to create more links in the future.

## References

Aadland, D., B. Anatchkova, B. D. Grandjean, J. F. Shogren, B. Simon, and P. A. Taylor. 2012. Valuing access to US public lands: A pricing experiment to inform federal policy. *Social Science Quarterly* 93 (1): 248–69.

Adamowicz, W., J. Louviere, and M. Williams. 1994. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management* 26 (3): 271–92.

Al-Ubaydli, O., M. S. Lee, J. A. List, C. L. Mackevicius, and D. Suskind. 2021. How can experiments play a greater role in public policy? Twelve proposals from an economic model of scaling. *Behavioural Public Policy* 5 (1): 2–49.

Allcott, H., and M. Greenstone. 2017. Measuring the welfare effects of residential energy efficiency programs. NBER Working Paper no. 23386, National Bureau of Economic Research, Cambridge, MA.

Alpizar, F., M. B. Del Carpio, and P. J. Ferraro. 2024. Input efficiency as a solution to externalities and resource scarcity: A randomized controlled trial. *Journal of the Association of Environmental and Resource Economists* 11 (1): 171–211.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman. 1993. Report of the NOAA Panel on Contingent Valuation. *Federal Register* 58: 4601–14.

Atkinson, G., B. Groom, N. Hanley, and S. Mourato. 2018. Environmental valuation and benefit-cost analysis in U.K. policy. *Journal of Benefit Cost Analysis* 9 (1): 97–119.

Australian Government. 2013. Clean Energy Act (Auction of Carbon Units) determination 2013. https://www.legislation.gov.au/F2013L00759/latest/text.

Baker, R., and B. Ruting. 2014. Environmental policy analysis: A guide to non-market valuation. Productivity Commission Staff Working Paper, Productivity Commission, Canberra, AU.

Balmford, B., J. Collins, B. Day, L. Lindsay, and J. Peacock. 2023. Pricing rules for PES auctions: Evidence from a natural experiment. *Journal of Environmental Economics and Management* 122: 102889.

Banzhaf, H. S. 2010. Economics at the fringe: Non-market valuation studies and their role in land use plans in the United States. *Journal of Environmental Management* 91: 592–602.

Bateman, I. J., I. H. Langford, R. Kerry Turner, K. G. Willis, and G. D. Garrod. 1995. Elicitation and truncation effects in contingent valuation studies. *Ecological Economics* 12 (2): 161–79.

Bateman, I. J., I. H. Langford, and J. Rasbash. 2001. Willingness-to-pay question format effects in contingent valuation studies. In *Valuing environmental preferences: Theory and practice of the contingent valuation method in the US, EU, and developing countries*, eds. Bateman, I. J., and Willis, K. G., 511–39. Oxford: Oxford University Press.

Benami, E., N. Jo, E. Ragnauth, and D. E. Ho. 2023. Drop a line, submit on time? Randomized tailored reminders improve pollution reporting timeliness. https://dx.doi.org/10.2139/ssrn.4355984.

Betz, R., K.-M. Ehrhart, B. Greiner, A. Ortmann, S. Schweitzer, and S. Seifert. 2010. Experimental testing of possible designs for the Australian carbon pollution permit allocation auction. Report to the Department of Climate Change and Energy Efficiency, Canberra, AU.

Betz, R., B. Greiner, S. Schweitzer, and S. Seifert. 2017. Auction format and auction sequence in multi-item multi-unit auctions: An experimental study. *Economic Journal* 127 (605): F351–F371.

BIT (Behavioural Insights Team). 2022. Increasing lead kit return rates among Chicago residents. Report, August 19. https://www.bi.team/wp-content/uploads/2023/09/2022-08-WWC-Chicago-DWM_Lead-Testing-Kits-Final-Report.pdf.

———. 2024. Leveraging behavioural insights to design and test the repairability index in France. Report, March 13. https://www.bi.team/publications/leveraging-behavioural-insights-to-design-and-test-the-repairability-index-in-france/.

Brown, C. H., G. Curran, L. A. Palinkas, G. A. Aarons, K. B. Wells, L. Jones, L. M. Collins, et al. 2017. An overview of research and evaluation designs for dissemination and implementation. *Annual Review of Public Health* 38: 1–22.

Browne, O. R., L. Gazze, M. Greenstone, and O. Rostapshova. 2025. Man vs. machine: Technological promise and political limits of automated regulation enforcement. *The Review of Economics and Statistics* 107 (4): 1136–48.

Burlig, F. 2021. Making energy efficiency work. In *U.S. Energy & Climate Roadmap: Evidence-based policies for effective action*, 110–21. Chicago: Energy Policy Institute at the University of Chicago.

Burtraw, D., J. Goeree, C. A. Holt, E. Myers, K. Palmer, and W. Shobe. 2009. Collusion in auctions for emission permits: An experimental analysis. *Journal of Policy Analysis and Management* 28 (4): 672–91.

Burtraw, D., C. A. Holt, K. Palmer, and W. Shobe. 2022. Price-responsive allowance supply in emissions markets. *Journal of the Association of Environmental and Resource Economists* 9 (5): 851–84.

California State Water Resources Control Board. 1994. *Decision 1631, Decision and order amending water right licenses to establish fishery protection flows in streams tributary to Mono Lake and to protect public trust resources at Mono Lake and the Mono Lake Basin*. Sacramento, CA: California State Water Resources Control Board.

Carlson, D., C. Forman, N. Olmstead, J. O. Ledyard, C. R. Plott, D. P. Porter, and A. Sholtz. 1993. An analysis and recommendation for the terms of the RECLAIM trading credit. Report to the South Coast Air Quality Management District, Diamond Bar, CA.

Carlsson, F., and O. Johansson-Stenman. 2012. Behavioral economics and environmental policy. *Annual Review of Resource Economics* 4 (1): 75–99.

Carson, R. T., and R. C. Mitchell. 1993. The value of clean water: The public's willingness to pay for boatable, fishable, and swimmable quality water. *Water Resources Research* 29 (7): 2445–54.

Cason, T. N. 1993. Seller incentive properties of EPA's emission trading auction. *Journal of Environmental Economics and Management* 25 (2): 177–95.

———. 1995. An experimental investigation of the seller incentives in the EPA's emission trading auction. *American Economic Review* 85 (4): 905–22.

Cason, T. N., and C. R. Plott. 1996. EPA's new emissions trading mechanism: A laboratory evaluation. *Journal of Environmental Economics and Management* 30 (2): 133–60.

Cason, T. N., L. Gangadharan, and C. Duke. 2003. A laboratory study of auctions for reducing non-point source pollution. *Journal of Environmental Economics and Management* 46 (3): 446–71.

Cason, T. N., and S. Y. Wu. 2019. Subject pools and deception in agricultural and resource economics experiments. *Environmental and Resource Economics* 73: 743–58.

CER (Commission for Energy Regulation). 2011a. Smart metering information paper 4: Results of electricity cost-benefit analysis, customer behaviour trials and technology trials. Information Paper no. CER11080. Commission for Energy Regulation, Dublin, IE.

———. 2011b. Smart metering information paper: Gas customer behaviour trial findings report. Information Paper no. CER11180a. Commission for Energy Regulation, Dublin, IE.

Connolly, D. J., G. Loewenstein, and N. Chater. 2025. An s-frame agenda for behavioral public policy research. *Behavioural Public Policy* 9 (3): 593–613.

Connor, J. D., J. Ward, C. Clifton, W. Proctor, and D. H. MacDonald. 2008. Designing, testing and implementing a trial dryland salinity credit trade scheme. *Ecological Economics* 67 (4): 574–88.

Cummings, R. G., D. S. Brookshire, and W. D. Schulze, eds. 1986. *Valuing environmental goods: An assessment of the contingent valuation method*. Totowa, NJ: Rowman and Allanheld.

Cummings, R. G., C. A. Holt, and S. K. Laury. 2004. Using laboratory experiments for policymaking: An example from the Georgia irrigation reduction auction. *Journal of Policy Analysis and Management* 23 (2): 341–63.

Deaton, A. 2020. Introduction: Randomization in the tropics revisited, a theme and eleven variations. In *Randomized control trials in the field of development: A critical perspective*, eds. Florent Bédécarrats, Isabelle Guérin, and François Roubaud, 29–46. Oxford: Oxford University Press.

Del Carpio, M. B., F. Alpizar, and P. J. Ferraro. 2021. Community-based monitoring to facilitate water management by local institutions in Costa Rica. *Proceedings of the National Academy of Sciences of the United States of America* 118 (29): e2015177118

DellaVigna, S., W. Kim, and E. Linos. 2024. Bottlenecks for evidence adoption. *Journal of Political Economy* 132 (8): 2748–89.

Duflo, E. 2018. The economist as a plumber. *American Economic Review* 107 (5): 1–26.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan. 2013. Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *The Quarterly Journal of Economics* 128 (4): 1499–545.

EPIC (Energy Policy Institute at the University of Chicago). 2021a. *U.S. Energy & Climate Roadmap: Evidence-based policies for effective action*. Chicago: Energy Policy Institute at the University of Chicago.

———. 2021b. Evidence-based policies to tackle today's energy and climate challenges. Insight / Impact Take-Away, December 13, 2021. https://epic.uchicago.edu/insights/evidence-based-policies-to-tackle-todays-energy-and-climate-challenges/.

Ferraro, P. J., T. L. Cherry, J. F. Shogren, C. Vossler, H. B. Flint, J. P. Hochard, O. Johansson-Stenman, et al. 2023. Create a culture of experimentation in environmental programs. *Science* 381 (6659): 735–37.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, et al. 2012. The Oregon health insurance experiment: Evidence from the first year. *The Quarterly Journal of Economics* 127 (3): 1057–106.

Fowlie, M., M. Greenstone, C. Wolfram. 2018. Do energy efficiency investments deliver? Evidence from the weatherization assistance program. *Quarterly Journal of Economics* 133 (3): 1597–644.

Foy, R., N. M. Ivers, J. M. Grimshaw, and P. M. Wilson. 2023. What is the role of randomised trials in implementation science? *Trials* 24 (1): 537.

Griffiths, C., H. Klemick, M. Massey, C. Moore, S. Newbold, D. Simpson, P. Walsh, and W. Wheeler. 2012. US Environmental Protection Agency valuation of surface water quality improvements. *Review of Environmental Economics and Policy* 6 (1): 130–46.

Hanley, N. 2026. Economic incentives for biodiversity conservation on farmland. *Review of Environmental Economics and Policy* 20 (1): 79–97.

Henrich, J., S. J. Heine, and A. Norenzayan. 2010. The weirdest people in the world. *Behavioral and Brain Sciences* 33: 1–23.

Herriges, J., C. Kling, C.-C. Liu, and J. Tobias. 2010. What are the consequences of consequentiality? *Journal of Environmental Economics and Management* 59 (1): 67–81.

Hjort, J., D. Moreira, G. Rao, and J. F. Santini. 2021. How research affects policy: Experimental evidence from 2,150 Brazilian municipalities. *American Economic Review* 111 (5): 1442–80.

Holt, C. A., and W. M. Shobe. 2016. Price and quantity collars for stabilizing emission allowance prices: Laboratory experiments on the EU ETS market stability reserve. *Journal of Environmental Economics and Management* 80: 69–86.

Holt, C., W. Shobe, D. Burtraw, K. Palmer, and J. Goeree. 2007. Auction design for selling CO2 emission allowances under the Regional Greenhouse Gas Initiative. Final report produced for RGGI, Resources for the Future, Washington, DC.

Johnston, R. J., K. J. Boyle, W. Adamowicz, J. Bennett, R. Brouwer, T. A. Cameron, W. M. Hanemann, et al. 2017. Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists* 4 (2): 319–405.

Jones & Stokes Associates. 1994. Final environmental impact report for the review of Mono Basin water rights of the city of Los Angeles. California State Water Resources Control Board, Sacramento, CA.

Krueger, A. B. 1999. Experimental estimates of education production functions. *The Quarterly Journal of Economics* 114 (2): 497–532.

Ledyard, J., D. Porter, and R. Wessen. 2000. A market-based mechanism for allocating space shuttle secondary payload priority. *Experimental Economics* 2: 173–95.

Lee, M.-Y., S. Steinback, and K. Wallmo. 2017. Applying a bioeconomic model to recreational fisheries management: Groundfish in the Northeast United States. *Marine Resource Economics* 32 (2): 191–216.

Levinson, A. 2016. How much energy do building codes save? Evidence from California houses. *American Economic Review* 106 (10): 2867–94.

List, J. A. 2024. Optimally generate policy-based evidence before scaling. *Nature* 626: 491–99.

List, J. A., and M. K. Price. 2016. The use of field experiments in environmental and resource economics. *Review of Environmental Economics and Policy* 10 (2): 206–25.

Loomis, J. B. 2000. Environmental valuation techniques in water resource decision making. *Journal of Water Resources Planning and Management* 126 (6): 339–44.

Metcalfe, P. J., W. Baker, K. Andrews, G. Atkinson, I. J. Bateman, S. Butler, R. T. Carson, et al. 2012. An assessment of the nonmarket benefits of the Water Framework Directive for households in England and Wales. *Water Resources Research* 48: W03526.

Moore, C. C., J. Corona, C. Griffiths, M. T. Heberling, J. A. Hewitt, D. A. Keiser, C. L. Kling, et al. 2023. Measuring the social benefits of water quality improvements to support regulatory objectives: Progress and future directions. *Proceedings of the National Academy of Sciences of the United States of America* 120 (18): e2120247120.

Narassimhan, E., K. S. Gallagher, S. Koester, and J. Rivera Alejo. 2018. Carbon pricing in practice: A review of existing emissions trading systems. *Climate Policy* 18 (8): 967–91.

NERA-Accent. 2007. The benefits of Water Framework Directive programmes of measures in England and Wales. Report CRP 4b/c, Department for Environment, Food and Rural Affairs, London, UK.

Normann, H.-T., and R. Ricciuti. 2009. Laboratory experiments for economic policy making. *Journal of Economic Surveys* 23 (3): 407–32.

Noussair, C. N., and D. P. van Soest. 2014. Economic experiments and environmental policy. *Annual Review of Resource Economics* 6: 319–37.

NRCB/CEAA (Natural Resources Conservation Board / Canadian Environmental Assessment Agency). 1998. Little Bow Project/Highwood Diversion Plan, Application to Construct a Water Management Project to Convey and Store Water Diverted from the Highwood River. NRCB/CEAA Joint Review Panel, Edmonton, AB.

OMB (Office of Management and Budget). 2023. *OMB Circular No. A-4*. Washington, DC: Executive Office of the President, Office of Management and Budget.

Petrolia, D. R., D. Guignet, J. Whitehead, C. Kent, C. Caulder, and K. Amon. 2021. Nonmarket valuation in the environmental protection agency's regulatory process. *Applied Economic Perspectives and Policy* 43 (3): 952–96.

Plott, C. R. 1997. Laboratory experimental testbeds: Application to the PCS auction. *Journal of Economics & Management Strategy* 6 (3): 605–38.

Porter, D., S. Rassenti, W. Shobe, V. Smith, and A. Winn. 2009. The design, testing and implementation of Virginia's NOx allowance auction. *Journal of Economic Behavior and Organization* 69 (2): 190–200.

Rolfe, J., J. McCosker, and J. Windle. 2008. Identifying the incentives that graziers in central-western Queensland need to manage land for conservation outcomes. *The Rangeland Journal* 30 (3): 297–303.

Roth, A. E. 1986. Laboratory experimentation in economics. *Economics & Philosophy* 2 (2): 245–73.

Schilizzi, S., and U. Latacz-Lohmann. 2007. Assessing the performance of conservation auctions: An experimental study. *Land Economics* 83 (4): 497–515.

Schmalensee, R., and R. N. Stavins. 2013. The $SO_2$ allowance trading system: The ironic history of a grand policy experiment. *Journal of Economic Perspectives* 27 (1): 103–22.

———. 2017. Lessons learned from three decades of experience with cap and trade. *Review of Environmental Economics and Policy* 11 (1): 59–79.

Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychological Science* 18 (5): 429–34.

———. 2018. The constructive, destructive, and reconstructive power of social norms: Reprise. *Perspectives on Psychological Science* 13 (2): 249–54.

Scott, J. T., K. M. Collier, J. Pugel, et al. 2023. SciComm optimizer for policy engagement: A randomized controlled trial of the SCOPE model on state legislators' research use in public discourse. *Implementation Science* 18: 12.

Shogren, J. F., J. Hochard, K. D. Lee, and L. V. Henderson. 2021. Experimental mindset for environmental challenges: The puzzling case of public good contributions. *European Review of Agricultural Economics* 48 (4): 785–804.

Shogren, J. F., and L. O. Taylor. 2008. On behavioral-environmental economics. *Review of Environmental Economics and Policy* 2 (1): 26–44.

Smith, V. L. 1994. Economics in the laboratory. *Journal of Economic Perspectives* 8 (1): 113–31.

Stoneham, G., V. Chaudhri, A. Ha, and L. Strappazzon. 2003. Auctions for conservation contracts: An empirical examination of Victoria's BushTender trial. *Australian Journal of Agricultural and Resource Economics* 47 (4): 477–500.

Stranlund, J. K. 2017. The economics of enforcing emissions markets. *Review of Environmental Economics and Policy* 11 (2): 227–46.

Thaler, R. H., and C. R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

US DOI (US Department of Interior). 1996. Record of decision: Operation of Glen Canyon Dam final environmental impact statement. Bureau of Reclamation, Salt Lake City, UT.

US EPA (US Environmental Protection Agency). 1996. Acid rain program, $SO_2$ allowance auction and electronic allowance transfer. *Federal Register* 61 (110): 28996–98.

US GAO (US General Accounting Office). 1994. Allowance trading offers an opportunity to reduce emissions at less cost. Report to the Chairman of the Environment, Energy, and Natural Resources Subcommittee, Committee on Government Operations. US House of Representatives, Washington, DC.

US SEC (US Securities and Exchange Commission), Office of Economic Analysis. 2007. Economic analysis of the short sale price restrictions under the regulation SHO pilot. https://www.sec.gov/news/studies/2007/regshopilot020607.pdf.

van Oirschot, P. 2015. Handhaven op afvalscheiding Terugblik na 1 jaar, handreiking voor anderen. Municipality of Tilburg, BAT department, December 2015. https://vang-hha.nl/publish/pages/112042/2016-05handhavenopafvalscheidingintilburgpvo.pdf.

Vollaard, B., and D. van Soest. 2024. Punishment to promote prosocial behavior: A field experiment. *Journal of Environmental Economics and Management* 124: 102899.

Vivalt, E., and A. Coville. 2023. How do policymakers update their beliefs? *Journal of Development Economics* 165: 103121.

Wallander, S., P. Ferraro, and N. Higgins. 2017. Addressing participant inattention in federal programs: A field experiment with the Conservation Reserve Program. *American Journal of Agricultural Economics* 99 (4): 914–31.

WaterSmart. 2014. Tapping into the power of behavioral science: Insights & opportunities for water-use efficiency. WaterSmart Software. https://www.watersmartsoftware.com/wp-content/uploads/2014/11/WSS_TappingintoBehaviorScience.pdf.

Welsh, M. P., R. C. Bishop, M. L. Phillips, and R. M. Baumgartner. 1995. Glen Canyon Dam, Colorado River storage project, Arizona—Non-use values final study summary report. NTIS No. PB98-106636. Hagler Bailly Consulting Inc., Madison, WI. National Technical Information Service, Springfield, VA.

Windle, J., J. Rolfe, J. McCosker, and A. Lingard. 2009. A conservation auction for landscape linkage in the southern Desert Uplands, Queensland. *The Rangeland Journal* 31 (1): 127–35.