

The role of experiments for policy design

Citation for published version (APA):

Werner, P., & Riedl, A. (2018). *The role of experiments for policy design*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 022
<https://doi.org/10.26481/umagsb.2018022>

Document status and date:

Published: 27/08/2018

DOI:

[10.26481/umagsb.2018022](https://doi.org/10.26481/umagsb.2018022)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Peter Werner, Arno Riedl

**The role of experiments for
policy design**

RM/18/022

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O. Box 616
NL- 6200 MD Maastricht
The Netherlands

The role of experiments for policy design

Peter Werner and Arno Riedl*

This version: July 2018

Abstract

In this paper, we review selected evidence to demonstrate the value of experiments for policy design with a focus on environmental policy and tax policy. Experiments can substantially improve ex-ante predictions about the outcomes of policy interventions, for example, by serving as “testbeds” to compare alternative market rules and mechanisms under tightly controlled conditions. Experiments also yield important insights into systematic deviations from strict rationality and into the heterogeneity of preferences among decision-makers that can form the basis for the (re-) design of policies. Besides describing various experimental approaches applied in the areas of environmental policy and tax policy, we also discuss further directions for successful collaborations between experimental economists and political decision-makers.

Keywords: Experiments, policy design, policy evaluation, behavioral regularities, non-standard preferences, environmental policy, tax policy

This is a draft article. The final version will be available in the “Handbook of Research Methods and Applications in Experimental Economics” edited by Arthur Schram and Aljaž Ule, forthcoming 2019, Edward Elgar Publishing Ltd.

The material cannot be used for any other purpose without further permission of the publisher, and is for private use only.

* Peter Werner, Department of Economics (AE1), School of Economics and Business, Maastricht University, P.O. Box 616, NL-6200 MD Maastricht, E-mail: p.werner at maastrichtuniversity.nl.
Arno Riedl, CESifo, IZA, Netspar, Department of Economics (AE1), School of Economics and Business, Maastricht University, P.O. Box 616, NL-6200 MD Maastricht, E-mail: a.riedl at maastrichtuniversity.nl.

1. Introduction

During the last decades, the importance of experiments in economic science has been steadily increasing and collaborations between experimental economists and policy makers are now established in a wide range of domains. Collaborating with policy makers can be a promising option for experimental researchers, because such collaborations offer the possibility to transfer insights from fundamental research to applied field problems with potentially substantial societal impact. The increasing importance of experiments for policy design is also reflected in the growing number of countries that have established (and are in the process of establishing) initiatives with the explicit goal of designing and evaluating public policy utilizing insights from behavioral economics. The Behavioral Insights Team in the UK founded in 2010 was the first policy unit among these institutions; in the Netherlands a Behavioral Policy Network has been established in 2014.¹

There are several reasons why the use of experiments can be an added value in the policy domain: First, experiments can improve the ex-ante assessment of outcomes of policy interventions. Policy measures are chosen to achieve a specific goal, and the nature of these measures reflects underlying theories about the mechanisms and channels through which this goal may be reached. Experiments can be used to validate these theories and also to compare alternative theories under controlled conditions before an intervention is implemented. For example, when the task is to design the rules in a specific market,

¹ The UNEP report on “Behavioural insights applied to policy” (Lourenço et al. 2016) gives an overview of European initiatives in this area. For an extensive list of recent interventions based on behavioral regularities please see the recent OECD report “Behavioural insights and public policy - Lessons from around the world” (OECD 2017).

experiments can serve as “testbeds” to compare the outcomes of alternative market institutions and therefore help to choose the option that fits the policy goal best.

Moreover, experimental techniques can improve the ex-post evaluation of policy measures. If a policy intervention is implemented in the field under a rigorous experimental protocol, observed changes in behavior among the target population can be unambiguously attributed to this intervention. A crucial feature of experiments is that they are based on the random assignment of persons into groups who are subject to variations in the experimental conditions (“treatments”) and a control group that serves as the reference condition. The comparison between the outcomes of interest in treatment groups and the control group avoids confounds due to selection problems present in non-experimental datasets (see, for example, Angrist and Pischke 2008 for a discussion of the value of random assignments). Experiments thus can establish causal effects between a policy measure and the related output variables and thereby add substantial precision to the ex-post evaluation of policy interventions.

Traditionally, economic policy measures have been designed from a neoclassical perspective based on the assumption that citizens are strictly rational (Chetty 2015). Yet experimental research both from the laboratory and the field has demonstrated that economic decision-makers are often only boundedly rational (Kahneman 2003) and exhibit other-regarding preferences such as concerns for fairness and reciprocity (Cooper and Kagel 2016). Also, people are found to be widely heterogeneous with respect to social preferences and other personal characteristics such as the attitude towards risk (see, for example, Bellemare et al. 2008, Dohmen et al. 2009, Dohmen et al. 2011, Falk et al. forthcoming). In this sense, fundamental research in experimental economics has

provided important insights into the nature of decision-making that can be utilized for the proper design of policy interventions.

The discrepancy between the assumption of full rationality and empirically relevant behavioral patterns might in many cases lead to significantly different assessments of the anticipated success of particular policy measures. Moreover, the inclusion of behavioral regularities in the decision process of politicians may lead to the emergence of new policies that are not identified by a neo-classical perspective, simply because the policy parameters would be decision-irrelevant if economic agents indeed would maximize their monetary utility (Madrian 2014, Chetty 2015).

In this paper, we provide a deliberately selective overview of diverse experimental methods in economics that can support political decision-making, covering both laboratory and field experiments. The advantages and caveats of these two kinds of experiments have been discussed extensively in the literature (see, for example, Harrison and List 2004, Levitt and List 2007, Falk and Heckman 2009, Levitt and List 2009, Croson and Gächter 2010). As we focus on the general role of experiments for policy design and evaluation, we will not compare the relative merits of each approach in this paper. From a policy as well as scientific perspective, both approaches should in our view be seen as complementary tools and be combined where necessary and possible.

The remainder of this paper is organized as follows: In the next section, we will review selected cases where experimental studies have created important insights for policy design. For the sake of brevity, we will concentrate on selected topics from the areas of environmental policy and tax policy. Of course, experiments have yielded important policy-relevant insights in many other areas (in the fields of health policy and the

regulation of retirement savings, to name just two). Also, as the scope of this paper does not allow us to discuss a larger number of thematically related studies, we will focus on exemplary papers with the goal to demonstrate the diversity of experimental approaches that can be applied in the policy domain. In the third and final section of this paper, we discuss some important factors that we believe will shape the collaboration between experimental economists and policy makers in the future.

2. Experiments and policy design - Applications

Environmental policy

Environmental protection is one of the most pressing problems for humanity. At the same time, its inherent social dilemma character makes it very difficult to mitigate the extensive exploitation of natural resources. Hence, a crucial question for policy makers in this context is how to induce stronger cooperation in the context of environmental protection. For instance, international cooperation in the context of the abatement of carbon dioxide emissions is seen as a key prerequisite to fight climate change (Cooper et al. 2016). Obviously, the climate change problem has many facets and involves activity on both collective and individual level, from negotiation between state governments to individual consumption decisions. In the following, we will give an overview of some dimensions of the climate change problem where experiments have provided important insights to support policy design.

First, targeted at private consumers, a growing number of field experiments test the impact of interventions designed to lower resource usage, often in cooperation with utilities (for example, Price 2014, Hahn and Metcalfe 2016, Brent et al. 2017 provide

extensive surveys in which they review many of the studies described here in more detail). A particular advantage of this approach is that experiments are conducted on a large scale and with “real” customers, allowing for direct inferences on the effectiveness of an intervention with the targeted population. The efficient use of energy is a cognitively challenging task and involves trading off short-term versus long-term benefits and costs, and it is unlikely that all consumers are able to perform the necessary calculations and to arrive at an individually optimal result. Moreover, incomplete information or inattention to the relevant decision parameters can substantially distort decisions.

In the context of energy use, various field experiments focus on customers’ responses to the introduction of dynamic pricing schemes. The goal of these schemes is to reduce electricity usage of the targeted customers in phases of peak demand by significant price increases during these periods. Standard economic reasoning would suggest that households adjust electricity use when facing higher prices. Yet, while the field evidence suggests that consumers do react to price shifts (see the surveys mentioned above for a summary of the evidence), the price elasticity of demand is on average smaller than expected and estimates vary greatly between the studies. This suggests that households might not react to price changes in a fully rational way or that they might lack the required information to adjust their energy consumption.

Indeed, non-market instruments explicitly targeted at potential information frictions induce significant shifts in energy conservation. For example, Jessoe and Rapson (2014) observe in a large field experiment that households indeed reduce energy consumption in response to significant price shifts, but to a significantly stronger extent if they have

adequate knowledge about their real-time consumption which is ensured by the provision of “in-home displays” allowing the treated households to track their energy consumption and electricity prices in real time. Moreover, Tiefenbeck et al. (2017) show in a sample of Swiss households that the availability of devices that provide real-time information about energy and water consumption while showering substantially reduces energy consumption and increases awareness about resource usage. At the same time, Allcott and Taubinsky (2015) test the importance of information provision about the cost savings potential of energy-efficient lightbulbs in an experiment on a survey platform with a representative population sample and in a natural field experiment and find only a moderate increase in the willingness to pay and no significant demand effect in the field. This finding suggests that inattention and lack of information alone cannot explain energy overconsumption.

Other field experiments have established an important impact of the provision of social information to customers, explicitly addressing non-standard preferences. One potential channel through which social information might change resource usage is through social norms. The underlying hypothesis is that communicating norm-relevant information to households might shift behavior into the direction of the norm due to a preference for norm conformity.

The importance of norms for electricity consumption is demonstrated by a field study conducted by Allcott (2011), drawing on a sample of 600,000 households in the United States. In a series of field experiments conducted in collaboration with local utilities, a random sample of households receives home energy reports from the company OPOWER containing comparison information about the electricity usage of neighbors. This

treatment is highly effective in achieving reductions of energy usage: Treated households reduce energy consumption by about 2% on average, with the largest effects found for households with high pre-treatment usage.² Also, Ferraro and Price (2013) consider the effects of one-time mailings including appeals to water conservation and find that appeals including social comparison information are especially effective in inducing lower water consumption although the general effect of appeals is getting weaker in the months after the intervention. Similarly, Allcott and Rogers (2014) observe a “backsliding” effect in the sense that households who receive OPOWER reports initially reduce energy consumption but this reduction is mitigated during the time period when no report is received. However, this backsliding effect becomes smaller with repeated receipts of the reports. Moreover, even in a group of households for which the intervention was stopped after two years, positive effects of the reports are found in the subsequent years, suggesting that the provision of energy reports including social information can have lasting behavioral impacts through habit formation.

The results of these studies are very appealing from a policy perspective, because they suggest that simple and cheap interventions can have noticeable effects on behavior. For example, Allcott (2011) estimates that the effect of the social comparison intervention on energy reduction is equivalent to the effect of an increase in the short-term energy price of about 10 - 20 %.

At the same time, customer responses to policy interventions might be very heterogeneous. For example, Costa and Kahn (2014) find that energy reports are much

² The treatment also included additional norm information (conveyed through emoticons related to the relative standing towards the peer group), but this information does not drive the main treatment effects.

more effective in inducing energy savings among political liberals than among conservatives. This and other findings (also in the studies described above) imply that there is a risk that information interventions might be ineffective for substantial shares of the targeted population due to the nature of the underlying preferences of these groups. Hence, when designing an intervention, policy makers have to gain more precise insights into the preference structure of the target population in order to tailor the details of the intervention to these preferences.

One effective way of generating such insights is the use of so-called “lab-in-the-field” experiments. Lab-in-the field experiments, defined by Gneezy and Imas (2017) as studies that are “conducted in a naturalistic environment targeting the theoretically relevant population but using a standardized, validated lab paradigm” enable researchers (and policy makers) to gain an understanding about the degree and nature of non-standard preferences among the target population.³

To gain a better understanding of behavioral factors relevant for the extraction of natural resources, several studies collect information about non-standard preferences on the individual level and link this to behavior in the field. For example, Rustagi et al. (2010) conduct lab-in the field experiments among forest user groups in Ethiopia who jointly manage the extraction of the resources from their forests. For individual forest users, the authors measure the inclination towards conditional cooperation (i.e. the willingness to cooperate if others also cooperate) and find that a higher share of conditional cooperators in a group is associated with more successful forest management. Moreover, additionally

³ Gneezy and Imas (2017) and Viceisza (2016) provide detailed surveys about lab-in-the-field experiments.

collected survey data reveals that these individuals also engage in more extensive monitoring activities and thereby enforce higher cooperation by other members of their groups.

Similarly, Fehr and Leibbrandt (2011) link measures for cooperativeness and time preferences of Brazilian fishermen with resource extraction in the field. They find that fishermen who are more cooperative and more patient in the experiment show less exploitative behavior in the field, as they use more sustainable fishing techniques. Riedl and Smeets (2017) show that such a relationship between preferences measured in a laboratory-like environment and field behavior not only holds for forest management and fishermen but extends to high stakes financial decisions. They find that social preferences measured in an Internet experiment using a trust game are predictive for the likelihood that private investors hold socially responsible mutual funds in their portfolio.

Several studies also investigate how the institutional and natural surroundings shape cooperativeness in the context of resource extraction. For instance, Gneezy et al. (2016) provide evidence that the working environment has a crucial impact on cooperativeness: A group of Brazilian fishermen who fish at sea and are forced to work in groups exhibit substantially higher cooperativeness in a range of experimental games than fishermen from a comparable society where fishermen work individually.

Experiments like the described studies allow for a deeper understanding of the distribution of preferences within a given population. In particular, as Gneezy and Imas (2016) point out, lab-in-the-field studies can yield important complementary insights for large scale field interventions and randomized control trials. Lab-in-the-field studies also allow politicians to gain understanding of the underlying motivations behind behavioral

patterns and to assess the prospects of success for policies that have the goal to trigger particular behavioral responses among citizens.

The research described above focuses on individual decision-makers. Other studies related to environmental policy address the role of the institutional environment. This line of research is devoted to the organization of market processes to internalize the externalities arising from the extraction of resources (see, for example, Sturm and Weimann 2006, Cason 2010, Friesen and Gangadharan 2013 and Noussair and van Soest 2014 for extensive surveys of the literature).

A prominent example where experiments have contributed to policy making in this area is the design of market institutions for the trading of emission permits, while recent studies also explore the trading of water and auctions with the goal to buy land for conservation purposes (see Friesen and Gangadharan 2013 for a comprehensive overview of this research). Within the last decades, large-scale markets for emission permits have been implemented in many countries including the US and notably the European Union with its large Emission Trading System (EU-ETS), with the goal to efficiently allocate pollution rights to emitters.

The rules and structure of these markets are typically very complex and it is often difficult to assess theoretically how adjustments in the market rules may affect the incentives for market participants and the resulting outcomes. Because field experiments are typically not possible and because failures in the design of large scale markets can lead to large economic losses and inefficiencies, here laboratory experiments can be an important tool in successful market design. Experimental studies in this area are often directed towards the inherent questions and problems of very specific markets. These

studies test predictions about market outcomes with human traders under tightly controlled conditions, enabling researchers to assess whether behavior converges to theoretical predictions and also to observe behavioral regularities that are not predicted by theory. Moreover, in cases where economic theory does not provide clear hypotheses, experiments have an important function as “testbeds” in order to evaluate the advantages and disadvantages of alternative market designs.

Given the large variety of auction designs and market parameters that these studies address, it is beyond the scope of this paper to provide a broad overview of the literature. Instead, we focus on outlining a few topics that have been in the focus of the experimental work on the design of emissions markets. First, seminal studies have focused on the role of the pricing rule in auctions for emissions rights. For example, Cason (1995) and Cason and Plott (1996) study the impact of the rules of an auction organized by the US Environmental Protection Agency to trade emission allowances and show how details in the pricing mechanism can substantially distort the auction outcomes. In this auction, bidders are matched to sellers according to bid prices, but with the specific feature that sellers who have posted the lowest asking price receive the highest bids. This rule creates undesired effects, as it becomes optimal for sellers to ask for a lower price than their true cost of abatement. Cason (1995) confirms this pattern in an experiment that covers only one market side (incentives are reversed in this design so that participants act as buyers who should overbid relative to their valuation). Cason and Plott (1996) test this auction design comparing a discriminatory pricing rule (every bidder pays the price according to his individual bid) to a uniform pricing rule (all bidders pay

the same price) and find that the latter achieves higher efficiency and leads to more truthful revelation of valuations and costs.

Another group of studies focuses on the impact of market rules concerning the expiration of emission certificates. For instance, Cason and Gangadharan (2006) test what the permission of “banking” - that is, the ability of emitters to keep unused certificates for future periods - has on price dynamics in these markets. Uncertainty for producers is implemented through random shocks to their planned abatement levels. The authors observe that banking reduces price variability, in particular after the realization of abatement shocks. At the same time, banking leads to lower compliance, suggesting that producers distort their reported abatement levels downwards in order to save certificates for later periods, which eventually leads to higher emissions.

As a final example, we emphasize studies that analyze the role the initial allocation of emission permits has for later market outcomes. In field emissions markets, certificates either have been assigned to the producers at no cost on the basis of historical emissions (“grandfathering”) or have been auctioned off. Theoretically, if producers can trade certificates after the allocation on a spot market, the initial allocation procedure should not matter for final outcomes. However, experiments show that behavior of bidders as well as market outcomes seem to respond to the initial allocation mechanism: For example, Goeree et al. (2010) find that in a treatment with grandfathering high emission producers exercise market power in the spot market and certificate prices increase above the competitive level. Moreover, as spot markets cannot fully correct the initial misallocations of certificates, low emission producers end up with too few certificates, in

particular in the grandfathering treatment, which lowers supply and decreases consumer welfare in a subsequent production stage.

Studies like the ones presented, but also more generally studies that test market designs often differ from “typical” laboratory experiments in important ways. In particular, such studies are often motivated by specific problems of the markets in question and are often aimed at directly informing policy makers about the impact of a-priori given alternative market institutions. Consequently, the experimental decision situation should resemble the relevant features of the particular market as closely as possible. Such studies, while being informative for the specific policy question, often have the downside of reduced generalizability.

Tax Policy

Given that taxes are the main source of income of the government, it is of vital importance to understand behavioral responses of citizens to the intended incentives of the existing tax system and to assess which interventions may increase tax revenues and compliance. Substantial experimental work has been conducted in two areas: The first group of studies focuses on tax evasion and investigates responses either to changes in the sanctioning mechanism or to interventions that aim at triggering the intrinsic motivation of citizens to pay taxes (Luttmer and Singhal 2014 provide an extensive survey). The second group of studies starts from the idea that citizens might not respond in a strictly rational way to the incentives set by a particular tax and tests if and how changes in the information or the presentation of taxes affect decisions.

Tax evasion is a widespread problem which Alm (2012, p.55) defines as “illegal and intentional actions taken by individuals to reduce their legally due tax obligations”. The standard economic approach to model the individual decision to evade taxes has been introduced by Allingham and Sandmo (1972). In their model, a rational taxpayer has to choose whether to report his true income or a smaller amount. Given the likelihood of tax audits and the potential fines imposed when caught, the decision-maker evaluates expected costs and benefits of underreporting taxable income. Allingham and Sandmo show that both an increase in the audit probability and the penalty after being detected lead to a higher reported income.

If the policy goal is to reduce tax evasion, it is necessary to assess the behavioral effects of policies aimed at increasing the likelihood of detecting tax evasion, as well as a thorough understanding of potential additional factors that drive compliance. Yet given the illegal nature of tax evasion, taxpayers have the obvious incentive to disguise their actions as much as possible, which makes it extremely hard to track individual evasion using field data - caught tax evaders, for which data are available, potentially represent only a small share of the actual cases. Here, the experimental approach has a substantial advantage over observational field data analysis, among other things because experiments can uncover the underlying mechanisms behind tax evasion and also enable researchers to test behavioral responses to changes in parameters of a tax regime.

Many early laboratory investigations related to tax policy have focused on the effect of audit probability and fines on the decision to evade taxes (Blackwell 2007, Alm and Jacobson 2007 and Alm 2012 review the literature) and generally found that these factors are positively correlated with tax compliance. Moreover, evidence from field experiments

suggest that the fear of being detected has a significant impact on individual decisions. For example, Kleven et al. (2011) conduct a large field experiment with a random sample of more than 40,000 Danish taxpayers. In the first part of this study, a group of taxpayers is randomly selected and subject to an audit process. The authors find evidence for substantial underreporting of self-reported income. In a second intervention conducted in the subsequent year, letters are sent out to different groups of taxpayers announcing different probabilities with which their tax statements will be audited. The authors observe that the threat of an audit increases self-reported income (with stronger effects when the audit is conducted with certainty). Notably, these effects are weaker than the effect of an actual experience of an audit in the previous year. A similar effect is found in the field experiment by Fellner et al. (2013) who analyze the impact of mailings to potential evaders of TV license fees and find among other things that receiving a letter significantly increases the probability of registering for the fee, and that this probability increases further if the mailing includes the threat of an audit.

These observations are important for policy design as they increase the confidence in the general effectiveness of measures to improve tax compliance under the assumption of rational taxpayers.

At the same time, in their survey article Luttmer and Singhal (2014) emphasize the importance of intrinsic motivation among citizens (“tax morale”) to comply with tax obligations. For example, Dwenger et al. (2016) conduct a field experiment in which the authors exploit the fact that the local church tax in the German federal state of Bavaria is not enforced. In a baseline treatment with no intervention, the authors find that despite

the lack of enforcement, a substantial share of taxpayers (more than 20%) indeed fulfill their church tax liability.⁴

This result opens up the possibility for policy makers to potentially induce higher tax compliance in the field by choosing adequate appeals to taxpayers, in order to leverage citizens' intrinsic motivation for compliance. Due to the low costs of implementation (especially compared to the costs of more extensive tax audits), an intervention targeting non-standard preferences appears very attractive, but at the same time, the design of such an intervention requires a thorough understanding of the underlying psychological mechanisms that drive tax compliance. Similar to the examples from energy policy, one potential mechanism is adherence to social norms. If norm conformity plays a role for tax compliance, approaching taxpayers with norm-relevant information might induce less evasion. Several field experiments have put this idea into practice, but the evidence so far is mixed. Among these studies, for instance, Bott et al. (2017) find that moral appeals have a significant positive effect on declared income from abroad. Moreover, moral appeals seem to be effective predominantly on the intensive margin, i.e. they induce higher payments from taxpayers who already declare some income. At the same time, the moral suasion intervention appears not to have a lasting effect as it does not influence declared income one year later.

Hallsworth et al. (2017) test a variety of messages that, for example, trigger different aspects of norms in a sample of more than 200,000 taxpayers in the United Kingdom and establish a positive effect of social norms messages on the timely payment of taxes.

Moreover, the field experiment by Perez-Truglia and Troiano (2015) shows that an

⁴ In the context of this particular church tax, citizens can even decide to pay more than their tax liability which is considered as a donation to the local church.

intervention that makes the tax delinquency of a person more salient to households who live nearby increases the probability of repaying taxes, but only if tax debts are small. Relatedly, Casal and Mittone (2016) find in a laboratory context that negative sanctions (showing photos of tax evaders) significantly increases compliance.

At the same time, there are examples in the literature where interventions related to payment norms were not successful. In the study by Fellner et al. (2013) described above, neither the information about the descriptive norm (the share of citizens who actually pay the tax) or a moral appeal highlighting the fairness aspect of paying the tax has a significant additional positive impact on compliance. Also, in a set of treatments conducted by Dwenger et al. (2016), announcing rewards for compliance (the possibility of public recognition or a monetary price) has a counterproductive effect as this reduces the probability of paying among previous tax evaders. In a laboratory setting, Lefebvre et al. (2015) show that norm conformity might backfire from a tax authority perspective: When participants receive information about evasive patterns in previous experimental sessions, evasion rates do not respond to information on high compliance, but increase in response to information on low compliance.

All in all, the literature leads to ambiguous conclusions about the impact of interventions that appeal to a preference for norm conformity. This ambiguity is important information for policy design as it shows that some policy instruments may only work in specific circumstances.

A second area in which experiments offer important insights is related to the perception of tax incentives. Overall, experimental studies from both the laboratory and the field suggest that the perception of taxes is substantially distorted (Fochmann et al. 2010

survey the earlier evidence), even in stylized settings where the incentives induced by the taxes are relatively easy to understand. The potential impact of distorted tax perceptions might be severe given that most actual tax systems are highly complex and non-linear.

First, there is abundant evidence for distortions in the understanding and evaluation of tax incentives. In addition, the way a tax is presented is also found to significantly affect behavior. For example, in a field experiment conducted in a grocery store, Chetty et al. (2009) find that increasing the saliency of a sales tax by incorporating it into the price tags significantly decreases the purchased quantities of the treated items. Blaufus et al. (2013) provide evidence from conjoint analysis and a laboratory study that the majority of experimental subjects, when being confronted with different menus of taxes, are unable to pick the menu of tax rate and tax base that would minimize their tax burden. Instead, decision-makers seem to rely on decision heuristics and overrate changes in the tax rates relative to changes in the tax base.

Second, an important related question is whether taxpayers are adjusting their labor supply to changes in the taxation of labor income which, as a result of various deductions possibilities relating to the personal situation of the taxpayer, might be rather challenging to understand. Since it is not possible in the field to vary taxes systematically for different groups of citizens, laboratory experiments can help to find answers to this question. In a real-effort experiment, Abeler and Jäger (2015) let subjects work under a simple and a complex tax regime that induce the same optimal output. The authors find that under the complex tax system, the probability that a subject chooses a suboptimal output level is higher. In addition, subjects who face the complex tax regime show weaker behavioral adjustments to the introduction of a new tax. Similarly, Blumkin et al. (2012) reject the

hypothesis that consumption and income taxes are behaviorally equivalent in a real effort experiment: When being confronted with a tax on labor income, subjects reduce their working time significantly stronger than subjects who have to pay an equivalent consumption tax.

The results obtained by Weber and Schram (2017) indicate that even the perception of which party legally pays a labor tax (employer or employee) can significantly affect subjects' responses. Although it should be irrelevant from a rational perspective which party legally pays the tax, employees express higher satisfaction levels and a preference for a larger public sector when the tax is levied at the employer's side, the latter being consistent with the interpretation that paying the tax themselves is perceived as a loss by employees. This violation of the liability side equivalence (LSE) of taxes (and subsidies) has also been found by Kerschbamer and Kirchsteiger (2000). In these studies taxes are imposed in respectively a multilateral and bilateral setting without markets. Other studies have explored LSE in competitive market settings and do not find a violation of LSE (Kachelmeier et al. 1994, Borck et al. 2002, Riedl and Tyran 2005, Ruffle 2005). This suggests that misperception of the legal and economic incidence of taxes plays a role in non-market settings but is absent or at least mitigated when taxes are imposed in competitive markets.

Besides the potential misperception of tax incentives, there is also evidence that the existence of taxes *per se* leads to additional distortions: In a real-effort experiment conducted with a non-student sample, Fochmann and Weimann (2013) find that subjects work significantly longer when their income is taxed compared to the case of an identical net wage without taxation. In a laboratory setting, Kessler and Norton (2016) report that

participants reduce working time significantly more in a real-effort task when lower remuneration is explicitly framed as a tax rather than as a wage reduction. Moreover, in a study that investigates how preferences for redistribution respond to different framings, Lorenz et al. (2017) find that participants choose substantially more redistribution when it is framed as a minimal income rather than as a tax.

All in all, the results of these studies clearly highlight that non-standard preferences may have a substantial impact on the willingness to fulfill tax obligations. Moreover, taxes which are equivalent from a rational perspective might induce non-trivial differences in behavior due to variations in their saliency and complexity. At the same time, the diverse findings in the literature so far do not allow for a systematic evaluation of the robustness and the relative importance of the documented observations.

One implication of the existing research is that due to the potential heterogeneity of taxpayers, simple interventions might not yield the desired outcomes. For example, Chetty and Saez (2013) show in a large field experiment that an information intervention that explains tax incentives to clients who receive earned income tax credits in the United States does not have a significant average effect on earnings in the subsequent year. Also, the mixed success of interventions highlighting norm-relevant information suggests that more emphasis should be put on exploring the heterogeneity of taxpayers. A recent field experiment by Boyer et al. (2016) moves into this direction: Similar to the study by Dwenger et al. (2016) discussed above, the study exploits the fact that a local church tax is not enforced. The authors send out different appeals to taxpayers that highlight the character of the tax either as legal obligation, voluntary payment or charitable donation and find that the responses to these appeals are substantially different between groups of

citizens who vary in their intrinsic motivation to pay the tax. Finally, only a few studies exist that compare behavioral responses from students to those of non-students. As these studies do not arrive at unambiguous conclusions (see, for example, Alm et al. 2015, Choo et al. 2016), more studies are required in order to determine to what extent the conclusions derived from laboratory settings with student participants are transferable to more representative samples.

3. Further directions of how experimental economists and political decision-makers can collaborate

The variety of the different experimental approaches reviewed in this paper demonstrates that the range of potential applications of experiments is large: First, studies conducted in the economic laboratory provide insights into fundamental behavioral responses to elements of policy interventions and serve as testbeds for alternative market institutions in the field. A good example for the latter application is the design of market institutions in the context of emission permits described above where actually existing institutions are analyzed with tools from economic theory and experiments with the goal of removing design flaws and improving market performance. As more and more market interactions are organized in a centralized and computerized way, we expect the importance of experiments providing “wind tunnel” tests for institutional rules to continue to be high (see also, for example, Roth 2002 for a description of the role of experiments in providing empirical evidence for market design, and Bolton and Ockenfels 2012 for an overview on the economic engineering of markets and institutions based on behavioral research).

Second, large-scale field experiments have established policy interventions based on research in behavioral economics and psychology as a powerful and cost-effective tool to influence behavior. Furthermore, a proper field study enables politicians to test the effectiveness of one or more interventions directly with the target population. More generally, guided by the prominence and success of large-scale field interventions in policy-relevant areas (Banerjee and Duflo 2016) we expect that the number of controlled field experiments will continue to grow, often to test the role of the exact details in the implementation that may be pivotal for the success or failure of policy measures (Duflo 2017).

From a researcher's perspective, a challenge for the collaboration between researchers and practitioners is the possible difference in objectives. From a scientific perspective, experimental research designs have to be as abstract as possible in order to remove potential confounds and to isolate the mechanisms underlying the results as cleanly as possible. Such experiments are likely too abstract and too general to be directly applicable to specific policy questions. However, some of the generated insights may still be policy-relevant: For example, experimentally detected violations of basic theoretical principles might caution policy makers about the limits of the theories on which they base decisions.

At the same time, experimental research designs that target specific policy design issues often have to include many details and specific content in order to be applicable to the policy issue in question. For example, experiments testing market designs (as discussed in the last section) or complicated economy-wide tax reforms (as in Riedl and van Winden 2007, 2012) tend to be more complex than typical laboratory studies. From a

practitioners' perspective it might be desirable to use experimental methods for targeted studies with a very narrow policy scope. Such studies might yield important insights that can be directly transferred to the field, but the generalizability of such insights would be limited (and potentially less of a concern from a practitioners' perspective).

Furthermore, there might be practical barriers for collaborations between experimental researchers and practitioners, in particular concerning the implementation of field interventions. Difficulties for practitioners to accept this method may arise from reservations against randomization (i.e. the fact that groups of individuals are randomly selected and treated while other groups do not receive the same treatment) or from concerns about legal issues including the active consent of individuals to participate in a field intervention. These concerns have to be weighted against the advantages that the random assignment of participants to treatments offer for the evaluation of policy measures (please see the introduction of this paper). Practitioners may also be concerned with the timing of experimental research due to the political cycle. While scientific scrutiny may require substantial time for the setup of the experimental design, practitioners may be more interested in a quick implementation.

One aspect that in our view has not always received sufficient attention is the long-term implementation of policy measures. In particular, in the context of large-scale field interventions, researchers should be involved in the long-term evaluation and the monitoring of the dynamic effects of one-time policy changes. A good example for the focus on the long-term performance of a policy intervention is the study by Allcott and Rogers (2014) who observe that repeated energy reports might help customers to develop

habits of lower energy consumption that create persistent energy savings even after the intervention has been stopped.

Moreover, heterogeneous responses are an important concern for the design of policy measures. If different citizen groups adjust their behaviors to an intervention in different ways, this might strongly mitigate the overall effect of a policy intervention. For example, heterogeneous responses might contribute to the explanation why appeals to non-standard preferences may not always induce higher tax compliance (see for example, the studies by Boyer et al. 2016 and Dwenger et al. 2016 who observe that the response to interventions depend on the characteristics of the particular taxpayers, and Costa and Kahn 2014 in the context of energy reports). Consequently, when citizens' preferences vary widely it may be unclear which kind of intervention will have the strongest behavioral impact. This calls for more research into the heterogeneity of preferences and how these are related to behavioral responses to policy interventions. Also, the exact behavioral mechanisms that may lead to specific responses to the interventions are often ambiguous. Hence, the design of interventions will typically have to incorporate multiple potential channels through which the chosen policy may affect behavior. To specify the exact design details and to avoid ambiguities as much as possible, ideally interventions should be based on theoretical hypotheses (see Card et al. 2011 for the value of theory for field experiments).

As policy interventions are often conducted in very specific domains, a natural question is to what extent results from these domains can be transferred to other settings. Given the complexity of field settings and the multiplicity of influence factors that may cause success or failure of a policy intervention, it is far from clear whether a measure that has

created positive effects in a certain application will have similar positive effects in another (similar) application. To assess the robustness and generalizability of particular policy measures to other domains, the successful implementation of an intervention should be replicated in similar environments and also in other policy areas. Such replications would be relevant both from a research and a policy perspective. This aspect connects to the more general issue of replication that is now prominently discussed in respect to laboratory experiments (see Camerer et al. 2016 and the references cited therein) but applies to all experimental (and non-experimental) methods. In many cases, replication might be problematic in the context of field experiments, as the exact conditions of the original study likely will be very difficult or impossible to replicate. This also holds for experiments with a focus on a very specific and narrow question of particular interest for practitioners. However, as some field interventions consist of very simple manipulations (for example, mailings that appeal to social preferences), the replication of the nature of these interventions in a related context should be in principle possible.

Relatedly, an important input for the design of effective policies is transparent information about experiments and interventions that have not induced the expected behavioral changes. This requires that experiments with null (or even negative) results are reported to a broad audience of scientists and political decision-makers. An example for the transparent communication of a null result is the study by Kettle et al. (2017) who conducted a large-scale field experiment with citizens from Guatemala who declare their taxes online. Before entering their tax filings, participants are presented information that should trigger different non-standard motivations to pay taxes, such as preferences for

honesty or the willingness to contribute to public goods. Yet, none of the interventions led to a significant increase in the declared tax. The authors attribute this finding among other things to the specific online setting in which the experiment was conducted and the self-selected sample of participants (individuals who made the active decision to declare their taxes and thus might not be affected by an intervention that aims at improving honesty). This highlights the fact that small details in the setup of a behavioral intervention can have a crucial impact on its effectiveness.

Studies in the laboratory enable researchers to isolate mechanisms that may drive patterns in the field but cannot be tracked there, and provide insights that can be used for the redesign of institutions in the field in turn. However, relatively few studies have exploited the advantages of different experimental methods in a complementary way. An optimal policy design cycle would comprise the complementary use of different experimental methods. It would start with a laboratory experiment to investigate the nature of preferences among human decision-makers that are relevant for the policy-relevant decision situation in the field. In case the existing experimental research already provides robust evidence on non-standard preferences, this first step might not always be necessary. The next step would be to conduct a lab-in-the-field study to gain thorough insights into whether the target group of a planned policy intervention has similar preferences and shows similar behavioral patterns as the laboratory population. Finally, a field experiment could be implemented to test behavioral responses of the target population to interventions designed on the basis of the insights into their preferences from the lab-in-the-field study.

In practice, however, such an approach may be difficult to implement due to financial and time constraints. A potential downside in that respect is that due to these limitations some interventions that are designed on the basis of less detailed knowledge about the preferences of the target population may be less successful than expected, which may have a negative effect on the perceived effectiveness of experiments for policy design in general. At the same time, the approach to use multiple experimental methods in a complementary way is in principle applicable to a wide area of policy-relevant problems and can provide important knowledge for the exact specification of policy measures. In particular, this approach enables targeted policies, i.e. policies designed for specific subgroups of the population that take the heterogeneity of preferences and decision-makers explicitly into account.

References

- Abeler, Johannes, and Simon Jäger. 2015. "Complex Tax Incentives." *American Economic Journal: Economic Policy* 7(3):1-28.
- Allcott, Hunt. 2011. "Social norms and energy conservation." *Journal of Public Economics* 95(9-10):1082-1095.
- Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* 104(10):3003-3037.
- Allcott, Hunt, and Dmitry Taubinsky. 2015. "Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market." *American Economic Review* 105(8):2501-38.
- Allingham, Michael G., and Agnar Sandmo. 1972. "Income tax evasion: a theoretical analysis." *Journal of Public Economics* 1(3-4):323-338.
- Alm, James. 2012. "Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies." *International Tax and Public Finance* 19(1):54-77.
- Alm, James, Kim M. Bloomquist, and Michael McKee. 2015. "On The External Validity Of Laboratory Tax Compliance Experiments." *Economic Inquiry* 53(2):1170-1186.
- Alm, James, and Sarah Jacobson. 2007. "Using Laboratory Experiments in Public Economics." *National Tax Journal* 60(1):129-152.
- Angrist, Joshua and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: an empiricist's companion*. Princeton, New Jersey: Princeton University Press.

Banerjee, Abhijit, and Esther Duflo. 2017. "Chapter 1 - An Introduction to the "Handbook of Field Experiments"." In *Handbook of Economic Field Experiments*, edited by Abhijit Vinayak Banerjee and Esther Duflo, 1-24. North-Holland.

Bellemare, Charles, Sabine Kröger, and Arthur Van Soest. 2008. "Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities." *Econometrica* 76(4):815-839.

Blackwell, Calvin. 2007. "A Meta-Analysis of Tax Compliance Experiments." International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University, Working Paper 07-24.

Blaufus, Kay, Jonathan Bob, Jochen Hundsdoerfer, Dirk Kieseewetter, and Joachim Weimann. 2013. "Decision heuristics and tax perception – An analysis of a tax-cut-cum-base-broadening policy." *Journal of Economic Psychology* 35:1-16.

Blumkin, Tomer, Bradley J. Ruffle, and Yosef Ganun. 2012. "Are income and consumption taxes ever really equivalent? Evidence from a real-effort experiment with real goods." *European Economic Review* 56(6):1200-1219.

Bolton, Gary E., and Axel Ockenfels. 2012. "Behavioral economic engineering." *Journal of Economic Psychology* 33(3):665-676.

Borck, Rainald, Dirk Engelmann, Wieland Müller, and Hans-Theo Normann. 2002. "Tax Liability-Side Equivalence in Experimental Posted-Offer Markets." *Southern Economic Journal* 68(3):672-682.

Bott, Kristina, Alexander Cappelen, Erik Ø. Sørensen, and Bertil Tungodden. 2017. "You've got mail: A randomised field experiment on tax evasion." NHH Dept. of

Economics Discussion Paper No. 10/2017.

Boyer, Pierre C., Nadja Dwenger, and Johannes Rincke. 2016. "Do norms on contribution behavior affect intrinsic motivation? Field-experimental evidence from Germany." *Journal of Public Economics* 144:140-153.

Brent, Daniel A., Lana Friesen, Lata Gangadharan, and Andreas Leibbrandt. 2017. "Behavioral Insights from Field Experiments in Environmental Economics." *International Review of Environmental and Resource Economics* 10(2):95-143.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu. 2016. "Evaluating replicability of laboratory experiments in economics." *Science* 351(6280):1433-1436.

Card, David, Stefano DellaVigna, and Ulrike Malmendier. 2011. "The Role of Theory in Field Experiments." *Journal of Economic Perspectives* 25(3):39-62.

Casal, Sandro, and Luigi Mittone. 2016. "Social esteem versus social stigma: The role of anonymity in an income reporting game." *Journal of Economic Behavior & Organization* 124:55-66.

Cason, Timothy N. 1995. "An Experimental Investigation of the Seller Incentives in the EPA's Emission Trading Auction." *American Economic Review* 85(4):905-922.

Cason, Timothy N. 2010. "What Can Laboratory Experiments Teach Us About Emissions Permit Market Design?" *Agricultural and Resource Economics Review* 39(2):151-161.

Cason, Timothy N., and Lata Gangadharan. 2006. "Emissions variability in tradable permit markets with imperfect enforcement and banking." *Journal of Economic Behavior & Organization* 61(2):199-216.

Cason, Timothy N., and Charles R. Plott. 1996. "EPA's New Emissions Trading Mechanism: A Laboratory Evaluation." *Journal of Environmental Economics and Management* 30(2):133-160.

Chetty, Raj. 2015. "Behavioral Economics and Public Policy: A Pragmatic Perspective." *American Economic Review* 105(5):1-33.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99(4):1145-77.

Chetty, Raj, and Emmanuel Saez. 2013. "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients." *American Economic Journal: Applied Economics* 5(1):1-31.

Choo, C. Y. Lawrence, Miguel A. Fonseca, and Gareth D. Myles. 2016. "Do students behave like real taxpayers in the lab? Evidence from a real effort tax compliance experiment." *Journal of Economic Behavior & Organization* 124:102-114.

Cooper, David and John Kagel. 2016. "Other Regarding Preferences: A Survey of Experimental Results." In *The Handbook of Experimental Economics, Vol. 2*, edited by John Kagel and Alvin Roth, 217-289. Princeton: Princeton University Press.

Cooper, Richard, Peter Cramton, Stéphane Dion, Ottmar Edenhofer, Christian Gollier, Éloi Laurent, David J.C. MacKay, William Nordhaus, Axel Ockenfels, Joseph Stiglitz, Steven Stoft, Jean Tirole, and Martin L. Weitzman. 2016. "Why Paris did not solve the

climate dilemma." In *Global Carbon Pricing - The Path to Climate Cooperation*, edited by Peter Cramton, David J.C. MacKay, Axel Ockenfels and Steven Stoff. Cambridge, Massachusetts: MIT Press.

Costa, Dora L., and Matthew E. Kahn. 2013. "Energy conservation "nudges" and environmentalist ideology: evidence from a randomized residential electricity field experiment." *Journal of the European Economic Association* 11(3):680-702.

Croson, Rachel, and Simon Gächter. 2010. "The science of experimental economics." *Journal of Economic Behavior & Organization* 73(1):122-131.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2009. "Homo Reciprocans: Survey Evidence on Behavioural Outcomes." *Economic Journal* 119(536):592-612.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9(3):522-550.

Duflo, Esther. 2017. "Richard T. Ely Lecture: The Economist as Plumber." *American Economic Review* 107(5):1-26.

Dwenger, Nadja, Henrik Kleven, Imran Rasul, and Johannes Rincke. 2016. "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany." *American Economic Journal: Economic Policy* 8(3):203-232.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. forthcoming. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics*.

Falk, Armin, and James J. Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science* 326(5952):535-538.

Fehr, Ernst, and Andreas Leibbrandt. 2011. "A field study on cooperativeness and impatience in the Tragedy of the Commons." *Journal of Public Economics* 95(9):1144-1155.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler. 2013. "Testing enforcement strategies in the field: threat, moral appeal and social information." *Journal of the European Economic Association* 11(3):634-660.

Ferraro, Paul J., and Michael K. Price. 2013. "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment." *The Review of Economics and Statistics* 95(1):64-73.

Fochmann, Martin, Dirk Kieseewetter, Kay Blaufus, Jochen Hundsdoerfer, and Joachim Weimann. 2010. "Tax Perception: An empirical survey." arqus Discussion Papers in Quantitative Tax Research No. 99.

Fochmann, Martin, and Joachim Weimann. 2013. "The Effects of Tax Salience and Tax Experience on Individual Work Efforts in a Framed Field Experiment." *FinanzArchiv: Public Finance Analysis* 69(4):511-542.

Friesen, Lana, and Lata Gangadharan. 2013. "Environmental Markets: What Do We Learn From The Lab?" *Journal of Economic Surveys* 27(3):515-535.

Gneezy, Uri, and Alex Imas. 2017. "Chapter 10 - Lab in the Field: Measuring Preferences in the Wild." In *Handbook of Economic Field Experiments*, edited by Abhijit Vinayak Banerjee and Esther Duflo, 439-464. North-Holland.

- Gneezy, Uri, Andreas Leibbrandt, and John A. List. 2016. "Ode to the Sea: Workplace Organizations and Norms of Cooperation." *The Economic Journal* 126(595):1856-1883.
- Goeree, Jacob K., Karen Palmer, Charles A. Holt, William Shobe, and Dallas Burtraw. 2010. "An experimental study of auctions versus grandfathering to assign pollution permits." *Journal of the European Economic Association* 8(2-3):514-525.
- Hahn, Robert, and Robert Metcalfe. 2016. "The Impact of Behavioral Science Experiments on Energy Policy." *Economics of Energy & Environmental Policy* 5(2):27-44.
- Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev. 2017. "The behavioralist as tax collector: Using natural field experiments to enhance tax compliance." *Journal of Public Economics* 148:14-31.
- Harrison, Glenn, and John List. 2004. "Field experiments." *Journal of Economic Literature* 42(4):1009-1055.
- Jessoe, Katrina, and David Rapson. 2014. "Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use." *American Economic Review* 104(4):1417-1438.
- Kachelmeier, Steven J., Stephen T. Limberg, and Michael S. Schade Wald. 1994. "Experimental Evidence of Market Reactions to New Consumption Taxes." *Contemporary Accounting Research* 10(2):505-545.
- Kahneman, Daniel. 2003. "Maps of bounded rationality: Psychology for behavioral economics." *American Economic Review* 93(5):1449-1475.
- Kerschbamer, Rudolf, and Georg Kirchsteiger. 2000. "Theoretically robust but

empirically invalid? An experimental investigation into tax equivalence." *Economic Theory* 16(3):719-734.

Kessler, Judd B., and Michael I. Norton. 2016. "Tax aversion in labor supply." *Journal of Economic Behavior & Organization* 124:15-28.

Kettle, Stewart, Marco Hernandez, Michael Sanders, Oliver Hauser, and Simon Ruda. 2017. "Failure to CAPTCHA Attention: Null Results from an Honesty Priming Experiment in Guatemala." *Behavioral Sciences* 7:28.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica* 79(3):651-692.

Lefebvre, Mathieu, Pierre Pestieau, Arno Riedl, and Marie-Claire Villeval. 2015. "Tax evasion and social information: an experiment in Belgium, France, and the Netherlands." *International Tax and Public Finance* 22(3):401-425.

Levitt, Steven, and John List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic Perspectives* 21(2):153-174.

Levitt, Steven, and John List. 2009. "Field experiments in economics: The past, the present, and the future." *European Economic Review* 53(1):1-18.

Lorenz, Jan, Fabian Paetzel, and Markus Tepe. 2017. "Just Don't Call it a Tax! Framing in an Experiment on Voting and Redistribution." *Journal of Experimental Political Science* 4(3):183-194.

Lourenço, Joana Sousa, Emanuele Ciriolo, Sara Rafael Almeida, and Xavier Troussard. 2016. *Behavioural insights applied to policy: European Report 2016*. European Union.

Luttmer, Erzo F. P., and Monica Singhal. 2014. "Tax Morale." *Journal of Economic Perspectives* 28(4):149-68.

Madrian, Brigitte. 2014. "Applying Insights from Behavioral Economics to Policy Design." *Annual Review of Economics* 6:663-688.

Noussair, Charles N., and Daan P. van Soest. 2014. "Economic Experiments and Environmental Policy: A Review." *Annual Review of Resource Economics* 6:319-337.

OECD. 2017. *Behavioural Insights and Public Policy: Lessons from Around the World*. OECD Publishing, Paris.

Perez-Truglia, Ricardo, and Ugo Troiano. 2015. "Shaming Tax Delinquents: Theory and Evidence from a Field Experiment in the United States." National Bureau of Economic Research Working Paper No. 21264.

Price, Michael K. 2014. "Using field experiments to address environmental externalities and resource scarcity: major lessons learned and new directions for future research." *Oxford Review of Economic Policy* 30(4):621-638.

Riedl, Arno, and Paul Smeets. 2017. "Why Do Investors Hold Socially Responsible Mutual Funds?" *Journal of Finance* 72:2505-2550.

Riedl, Arno, and Jean-Robert Tyran. 2005. "Tax liability side equivalence in gift-exchange labor markets." *Journal of Public Economics* 89(11):2369-2382.

Riedl, Arno, and Frans van Winden. 2007. "An experimental investigation of wage

taxation and unemployment in closed and open economies." *European Economic Review* 51(4):871-900.

Riedl, Arno, and Frans van Winden. 2012. "Input versus output taxation in an experimental international economy." *European Economic Review* 56(2):216-232.

Roth, Alvin E. 2002. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica* 70(4):1341-1378.

Ruffle, Bradley J. 2005. "Tax and subsidy incidence equivalence theories: experimental evidence from competitive markets." *Journal of Public Economics* 89(8):1519-1542.

Rustagi, Devesh, Stefanie Engel, and Michael Kosfeld. 2010. "Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management." *Science* 330(6006):961-965.

Sturm, Bodo, and Joachim Weimann. 2006. "Experiments in Environmental Economics and Some Close Relatives." *Journal of Economic Surveys* 20(3):419-457.

Tiefenbeck, Verena, Lorenz Götte, Kathrin Degen, Vojkan Tasic, Elgar Fleisch, Rafael Lalive, and Thorsten Staake. 2017. "Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation." *Management Science* 64(3):1458 - 1476.

Viceisza, Angelino C. G. 2016. "Creating a Lab in the Field: Economics Experiments for Policymaking." *Journal of Economic Surveys* 30(5):835-854.

Weber, Matthias, and Arthur Schram. 2017. "The Non-equivalence of Labour Market Taxes: A Real-effort Experiment." *The Economic Journal* 127(604):2187-2215.