American Economic Review 2021, 111(5): 1442–1480
https://doi.org/10.1257/aer.20190830

How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities[†]

By Jonas Hjort, Diana Moreira, Gautam Rao, and Juan Francisco Santini*

Can research findings change political leaders' beliefs and policies? We use experiments with 2,150 Brazilian municipalities to measure mayors' demand for and response to research information. In one experiment, we find that mayors are willing to pay to learn the results of evaluation studies, and update their beliefs when informed of the findings. They value larger-sample studies more, while not distinguishing between studies in rich and poor countries. In a second experiment, we find that informing mayors about research on a simple and effective policy, taxpayer reminder letters, increases the probability the policy is implemented by 10 percentage points. (*JEL* D72, D78, D83, O17, O18)

Recent decades have seen an explosion of program evaluation research in economics.[1] But how interested in and open to academic research *are* political leaders? And, insofar as they "consume" research, can and do they act on new findings? These are questions of fundamental importance for the science ecosystem. Despite the money and effort devoted to evaluating policy impact, we have little understanding of whether the conditions necessary for the public to ultimately benefit hold: whether political leaders *value* such research; whether it *changes their beliefs* about policy effectiveness; and whether leaders ultimately *implement* policies that they

* Hjort: Columbia University, BREAD, CEPR, and NBER (email: hjort@columbia.edu); Moreira: University of California at Davis (email: dsmoreira@ucdavis.edu); Rao: Harvard University, NBER, and BREAD (email: grao@fas.harvard.edu); Santini: Innovations for Poverty Action (email: jsantini@poverty-action.org). Esther Duflo was the coeditor for this article. We thank the mayors and municipal officials who gave us their time; Paulo Ziulkoski, Glademir Aroldi, Gustavo Cezario, Tatiane de Jesus, Jasmim Madueno, Zione Rego, and others at CNM for their enthusiastic collaboration; Lemann Brazil Research Fund at Harvard, the Weiss Family Fund, the Warburg Fund, and a JPAL Governance Initiative Pilot Grant for financial support; Teresita Cruz Vital, Xinyue Lin, Brian Wheaton, Felipe Lima, and especially Deivis Angeli and Vinícius Schuabb for excellent research assistance; three anonymous referees, Alberto Alesina, Juliano Assunção, Doug Bernheim, Nicolas Caramp, Stefano DellaVigna, Pascaline Dupas, Benjamin Enke, Bruno Ferman, Claudio Ferraz, Gustavo Gonzaga, Gianmarco León, Joana Naritomi, Muriel Niederle, Rohini Pande, Ricardo Perez-Truglia, Rudi Rocha, Chris Roth and seminar audiences at Columbia, the FGV-EESP, Harvard, PACDEV, PUC-Rio, Stanford, UC Davis, and USC for helpful suggestions. This project underwent ethics review by the Committee on the Use of Human Subjects at Harvard University. The two experiments reported in this paper are registered at the AEA Random Control Trial Registry, numbers AEARCTR-0004274 and AEARCTR-0004273. The initial draft of this paper was part of Juan Santini's PhD. dissertation at PUC-Rio.

[†] Go to https://doi.org/10.1257/aer.20190830 to visit the article page for additional materials and author disclosure statements.

[1] For example, more than 2,500 studies have been registered with the American Economic Association's registry for randomized controlled trials (RCTs) since its launch in May 2013.

otherwise would not have in response to new research findings. In short, is a lack of (access to) research information a binding constraint on policy choice?

In this paper, we take a first step toward answering these questions by providing evidence from two experiments. We leverage an unusual collaboration with the National Confederation of Municipalities (*Confederação Nacional de Municípios*, or CNM) in Brazil. We first report results from a beliefs experiment measuring policymakers' willingness-to-pay (WTP) to learn the findings of rigorous impact evaluations, as well as how such findings affect their beliefs. A total of 764 municipal officials (primarily mayors) from 579 municipalities participated in this first experiment. To estimate the ultimate impact on actual policy adoption, we use a second, larger-scale policy-adoption field experiment with 1,818 Brazilian mayors. A randomly selected treatment group of mayors was invited to attend a research-information session at a large CNM convention. A presenter informed the audience about the findings of a set of RCTs showing positive effects of a taxpayer reminder letter policy on tax compliance. We then measured not just beliefs about policy effectiveness, but the actual use of such reminder letters at the municipality level 15 to 24 months later. In combination, the beliefs and policy-adoption experiments allow us to estimate both the extent to which research findings influence policy if directly provided to political leaders, and the intermediate steps of policymaker demand and belief change that are one pathway through which research may impact policy.

Brazil's municipalities are an excellent setting to investigate how research affects policy practice for two reasons. First, their political leaders hold a role analogous to that of many countries' head of state: Brazilian mayors are directly elected and individually wield considerable de jure power over policy choices within the areas that municipalities' control.[2] Second, there are 5,570 municipalities in Brazil, and our collaboration with CNM gives us direct access to their leadership. CNM is a nonpartisan organization whose membership comprises thousands of municipal governments. It seeks to provide training and technical support to mayors and municipal managers, and advocates for their interests at the federal level. Working with CNM allowed us to carry out experiments at the polity level.

Beliefs Experiment.—Our first experiment finds that the political leaders of Brazil's municipalities exhibit significant personal demand for research and change their beliefs in response to research findings. The policy context for the experiment is Early Childhood Development (ECD) programs, whose impacts on children's test scores have been estimated in existing research. We make use of four comparable RCTs conducted in different locations and with different sample sizes.[3] Our experiment begins by eliciting beliefs about the likely impact of an ECD program if implemented in the participant's own municipality. We then present the participant with one randomly selected study, mentioning two study characteristics (location and

[2] In Brazil, municipalities control policy areas such as preschool and primary education, and preventative health and sanitation. Over 90 percent of Brazilian municipalities raise tax revenues locally, primarily from property and service taxes, in addition to the federal and state transfers they receive.

[3] The studies we use are Grantham-McGregor et al. (1991), Walker et al. (2005), Puma et al. (2010), Schweinhart et al. (2005), Barnett (2011), and Attanasio et al. (2014). These are all high-quality studies of the impact of ECD in, respectively, Jamaica (first two studies), the United States as a whole, Michigan, and Colombia, with varying sample sizes.

sample size). We elicit the participant's personal WTP to learn the study's results using an incentive-compatible procedure, and then randomize whether the individual actually receives the result (conditional on their WTP).[4] To deal with selection into receiving the study findings, a subset of participants receive the findings for free. If the results of the study are revealed, we elicit the participant's posterior beliefs about the likely effect of the policy. We also elicit incentivized beliefs about the likely effect in the contexts where the policy was actually implemented and evaluated. Finally, we offer the participant the opportunity to pay for practical advice on how to implement the ECD program. The entire experiment was self-administered privately by the participant using a tablet.

We find that while participants hold widely varying beliefs about the impact of the ECD policy to begin with, they are willing to pay an arguably fairly high amount (out of an experimental budget) to find out the results of an impact evaluation: about US$36 on average (under certain assumptions to benchmark the experimental currency). The average WTP is higher for studies with a large sample size but *not* for studies conducted in a location that is closer to Brazil's income level.[5] Learning the results of an RCT causes officials to update their beliefs about impact: their posterior is a weighted average of their prior and the revealed study's findings. Consistent with the demand (WTP) findings, policymakers update their beliefs more when they receive large-sample studies, but not when they receive studies conducted in developing countries rather than the United States. While we cannot rule out that these different responses to different studies are in part driven by attributes that participants expect to correlate with sample size and study location, the two study characteristics we explicitly state, 59 percent of the participants who report preferring the large-sample studies in a debriefing survey report statistical precision as a reason.

Our experiment is not designed to test a model of rational learning. Since we do not measure probabilistic beliefs and only provide participants with point estimates from the studies, we do not know how much participants *should* update their beliefs. We do, however, provide suggestive calculations that policymaker sensitivity to sample size is lower than a Bayesian model would predict. We also document substantial heterogeneity in how much individual policymakers update in response to information, including about one-quarter who do not update at all. Although limited in statistical precision, we find little evidence for two types of motivated reasoning. Specifically, participants on average do not display confirmation bias, they do not interpret information in a way that tends to reinforce their prior and lead to polarization, nor do they respond asymmetrically to good versus bad news regarding the policy (relative to their prior).

Altogether, the findings of our first experiment suggest that on average political leaders value research information and place substantial weight on it, at least once such information is made (easily) accessible. In line with this interpretation, we find that a higher posterior causally increases the policymaker's WTP for practical information on how to implement the policy. However, important caveats

[4] WTP is elicited in terms of an experimental currency. Specifically, each participant is endowed with lottery tickets with a chance to win an expenses-paid trip to the United States. They may instead use some of these lottery tickets to purchase access to the findings of the research.

[5] Limited by the existing evidence base, we could not offer participants a similar study from Brazil itself. It could well be that participants *would* be willing to pay more for such a study.

apply. First, the study measures WTP out of the policymaker's private experimental budget, rather than out of a municipal budget. It thus captures what the policymaker is personally willing to give up to acquire such information, but not whether they would be willing or able to spend out of government budgets, which may have other, higher-value uses or binding restrictions. Second, it only measures very short-run effects on beliefs. Third, the WTP and overall responsiveness to information may be affected by experimenter demand effects, even though the experiment was privately self-administered on a tablet. The following experiment partially addresses these weaknesses by studying actual policy adoption, a higher-stakes and longer-run outcome.

Policy-Adoption Experiment.—In our second experiment, we invited a randomly chosen subset of the mayors attending CNM's 2016 *Novos Gestores* convention in Brasília, the heads of 1,818 municipal governments, to attend an optional research-information session.[6] The policy tool discussed in the session was reminder letters to taxpayers to induce them to comply with taxes. We chose this policy both because its impact is well documented in existing, rigorous research, and because it is inexpensive and easy to implement. During the 45-minute long information session, an experienced local presenter introduced the idea of impact evaluation, described taxpayer reminder letters and their content, and presented research findings from studies on the quantitative impact of such letters on tax compliance.[7] At the end of the session, mayors were provided with a printed policy brief summarizing the information.

Of the randomly invited mayors in the treatment group, 37.9 percent chose to attend the information session. This is arguably a fairly high attendance rate, given that contact information was out-of-date for some mayors, and considering the meaningful opportunity cost: professional networking with other politicians, or attending parallel sessions on other topics which did not emphasize research findings. Younger and college-educated mayors were more likely to attend, while term-limited mayors were no less likely to attend than mayors in their first term.

Attending the research-information session increased the probability that municipalities had implemented taxpayer reminders 15–24 months later by 10 percentage points, or 33 percent relative to the 32 percent of municipalities in the control group which already implemented the policy.[8] There is little evidence of heterogeneity in treatment effects by leader or municipality characteristics; for example, term-limited mayors appear to be equally likely to attend the information session and to adopt reminder letters as mayors who face reelection incentives.

[6] The sampling frame consists of Brazilian municipalities with populations between 5,000 and 100,000 inhabitants for which the mayor was confirmed to attend the *Novos Gestores* convention. Forty-five percent of all mayoral administrations in Brazil within the relevant population range went to Brasília and thus were part of our sample. There are 881 municipalities in the treatment group and 937 municipalities in the control group.

[7] The findings that were presented at the information session were based on the following studies: Coleman (1996); Hasseldine et al. (2007); Del Carpio (2013); Fellner, Sausgruber, and Traxler (2013); Castro and Scartascini (2015); and Hallsworth et al. (2017).

[8] We surveyed key bureaucrats in treatment and control municipalities with knowledge of the municipality's tax policies (typically in the finance department) from February to November 2018, 15–24 months after the *Novos Gestores* convention, to verify whether taxpayer reminder letters were being implemented in the municipality. In 81 percent of the municipalities in the sample, at least one public official was surveyed. There was no differential attrition between treatment and control municipalities.

We interpret the effects on policy adoption as being driven by the provision of research information on policy effectiveness. Consistent with this, we find persistent effects on beliefs about the effectiveness of reminder letters, and evidence that beliefs change not just for the treated mayors, but also among their tax bureaucrats. However, a number of alternative channels and interpretations are important to acknowledge. First, it could be that adoption occurred simply due to learning of the existence of tax reminders. While this is possible, recall that tax reminders are not entirely unknown: one-third of the control group already uses them. Second, it could be that the intervention simply raised the salience of tax compliance as a policy goal, leading to the adoption of an already-known policy of reminder letters. Here, it is worth noting that we do not find any effects on the adoption of another commonly used tax policy, financial incentives for compliance. Third, it could be that some other effective policy was crowded out by the adoption of the reminder letters. We discuss this issue in more detail in Section III, but note that it does not contradict the conclusion that providing research information changed policy. Finally, it could be that a direct policy recommendation from the experimenters would have similar effects, even shorn of any underlying evidence.

The two experiments have similar structures but different strengths. In the policy-adoption experiment, mayors must pay for the information with their time, belief changes are measured over 15–24 months, we capture belief spillovers to local bureaucrats, and actual policy adoption is observed, albeit for a simple and low-cost policy. But this experiment does not shed light on what type of research information is more or less compelling, nor does it allow us to rigorously study belief-updating due to an inability to measure prior beliefs. In contrast, the beliefs experiment studies belief changes over only a matter of minutes. But it allows us to learn that policymakers respond to studies differently based on sample size but not location, and to shed light on heterogeneity in belief updating as well as explore deviations from Bayesian learning. While the magnitudes of effects are difficult to compare between the two complementary studies, the findings are qualitatively consistent. Policymakers are interested in research information; it changes their beliefs; and these changed beliefs can translate into policy change.

Numerous open questions remain. The two experiments studied different policies, which introduces a gap in our argument. Future efforts might measure policymaker beliefs and information-demand across more policy topics, and examine whether information is a binding constraint not just for the adoption of inexpensive and simple policies such as reminder letters, but also for more challenging and expensive (but effective) policies such as ECD programs. Understanding the credibility of different information sources is another important question for research. In this project, a trusted partner organization and researchers from reputed universities (Columbia, Harvard, and PUC-Rio) organized an information session. Other sources through which research information is encountered, such as local think tanks, academics or media sources, may be received differently. This paper also does not capture the numerous less-direct channels through which research may influence policy, such as by gradually changing ways of thinking, influencing donors and other non-state actors, or informing citizens. Finally, if policymakers do value research information and react to it, as we argue, this raises an important question: what prevents them from acquiring such information already? In the absence of direct outreach from

researchers, as in our project, how do policymakers discover and parse research findings? We hope that future work will shed light on these questions.

This paper contributes to and bridges the literatures on state effectiveness on the one hand, and the role of evidence and experts' beliefs on the other. The former has focused on selection into the state enterprise, and variation in politicians' and public sector workers' effectiveness under different incentive schemes.[9] Using a polity-level field experiment somewhat parallel to the management interventions in private firms studied in Bloom et al. (2013), we instead show that information frictions at the top, i.e., heads of government's lack of knowledge of policies' effectiveness, directly constrain policy decisions.[10] Our findings make clear that it is not the case, for example, that counterfactual policies' effectiveness is widely known "on the ground," nor that political leaders are uninterested in, unconvinced by, or unable to act on new research information. This implies that policy research *can* help political leaders improve their constituents' lives.

By starting to unpack how political leaders' beliefs are shaped, and their consequences, we also advance an emerging body of evidence on belief formation and the role of evidence. While most such research studies beliefs in lay populations to identify systematic biases and heuristics (see Benjamin 2019 for a review), we add to the smaller body of work studying the beliefs of experts such as central bankers (Malmendier, Nagel, and Yan 2017), academics (DellaVigna and Pope 2018), and judges (Chen, Moskowitz, and Shue 2016).[11] In this sense, our study is most closely related to Banuri, Dercon, and Gauri (2019); Nellis et al. (2019); Rogger and Somani (2019); and Vivalt and Coville (2020), who study how the beliefs of policy professionals, program officers, aid-agency workers, and bureaucrats respond to research findings and new data. Like these papers, we document substantial belief updating among policymakers in response to providing objective evidence. Our main contributions relative to those papers are to study heads of government, to shed light on the kinds of studies such policymakers value and place more weight on, to measure demand for such information, and, most importantly, to provide evidence that research evidence actually translates into changes in policies adopted.[12] It also complements recent research showing that citizens do change

[9] The literature on state effectiveness often views states as organizations and has focused on frontline public sector workers (see Finan, Olken, and Pande 2017 for a review), bureaucrats (see, e.g., Duflo et al. 2013, 2018; Nath 2015; Khan, Khwaja, and Olken 2016, 2019; Akhtari, Moreira, and Trucco 2018; Bertrand et al. 2020; Best, Hjort, and Szakonyi 2018; Rasul and Rogger 2018, among others), and leaders' identities (Chattopadhyay and Duflo 2004; Jones and Olken 2005; Besley, Montalvo, and Reynal-Querol 2011; Beaman et al. 2012; Martinez-Bravo 2014; Yao and Zhang 2015; Easterly and Pennings 2017; Martinez-Bravo 2017; Bertrand et al. 2020; Xu 2018). For an overview of the literature on politician motives, see Persson and Tabellini (2002).

[10] In this sense the existing study closest to ours is Hoffmann et al. (2017). They carry out an innovative lab-in-the-field incentive-compatible choice experiment in which elected county councilors in Kenya chose among alternative water infrastructure projects. Other influential polity-level natural and field experiments such as Fujiwara and Wantchekon (2013) and Bidwell, Casey, and Glennerster (2020), and related studies in political science, have randomized how electoral campaigns take place across electoral districts or villages and studied the impact on electoral outcomes.

[11] Our policy-adoption experiment builds on the influential information-provision approach pioneered by Jensen (2010) and many related studies (see, among others, Kling et al. 2012, Chetty and Saez 2013, Dizon-Ross 2019).

[12] Banuri, Dercon, and Gauri (2019); Nellis et al. (2019); and Vivalt and Coville (2020) study the belief-formation of (mostly UK- and US-based) policy professionals, while Rogger and Somani (2019) study the beliefs of bureaucrats in Ethiopia. Like Vivalt and Coville (2020), we find some evidence of precision-neglect, but unlike them, we do not find evidence of asymmetric responses to positive and negative news. Like Rogger and Somani (2019), we find little heterogeneity in belief-updating by policymaker or municipality characteristics. They emphasize

their policy preferences in response to evidence, even on controversial topics such as immigration (Grigorieff, Roth, and Ubfal 2020; Haaland and Roth 2020).

The rest of the paper proceeds as follows. Section I provides institutional information about Brazilian local governments and our partner organization. Section II presents the design and results from the beliefs experiment. Section III discusses our second intervention, the policy-adoption experiment, and finally we conclude in Section IV.

I. Institutional Background and Context

This section provides relevant background information on municipal governments in Brazil, our partner organization, and the conferences where our experiments were conducted.

A. *Brazilian Municipalities*

Municipalities are the lowest level of government in Brazil. In total, there are 5,570 municipalities distributed across 26 states. Municipal governments are headed by elected mayors, who appoint secretaries to lead the municipal bureaucracy. Once elected, mayors serve a four-year term and can hold office up to two consecutive terms. Elections are generally considered fair, such that politicians face some electoral accountability.

In Brazil, as in many Latin American countries, provision of services is generally devolved to municipalities, while revenue generation and collection is partially devolved. Municipal governments are responsible for key public services such as education, health, sanitation, and transportation. To cover the costs, municipalities rely in part on intergovernmental transfers. On average, 60 percent of municipalities' total revenues are transfers from state governments and the federal government. Part of the remainder is locally raised by municipalities themselves. Municipal governments are responsible for collecting local taxes, which represent on average 15 percent of municipal revenues.

In general, municipal governments are highly autonomous. The mayor negotiates the budget allocation with the city councilors and has full autonomy over its execution. The mayor's office thus holds policymaking authority over a wide range of areas. Our research information experiments will involve two such areas: early-childhood education and locally raised taxes. We describe these two areas in more detail in Sections II and III.

B. *Our Partner Organization*

This study leveraged a unique opportunity to conduct a series of large-scale experiments with thousands of local political leaders through a partnership with Brazil's National Confederation of Municipalities (CNM). CNM is a nonpartisan organization that serves as a coordinating body and advocate of Brazilian

heterogeneity by organizational management practices, which we are unable to observe. Another related paper is Beynon et al. (2012), which uses an online experiment to study the optimal design of policy briefs.

municipalities' interests at the state and federal level. Over 80 percent of all Brazilian municipalities are members of CNM. Importantly for our purposes, CNM organizes a variety of conferences and conventions throughout the year, in which thousands of municipal officials from all over the country participate.

These meetings provide an unusual opportunity to reach a large population of political leaders in one place. Meeting attendees comprise mayors, vice-mayors, local legislators, and municipal secretaries. Our beliefs experiment was conducted at two of CNM's annual national conventions (May 2017 and May 2018) and at 12 regional conferences held in different states (August–December 2017).[13] Our policy-adoption experiment was conducted at CNM's biggest national conference, called *Novos Gestores*, which is held every four years in Brasília (October–November 2016). All mayors who were (re)elected in the last municipal election are invited to attend *Novos Gestores*.

Our research-information interventions were one of the many activities that took place at these meetings. The meetings are each two or three days long, and are structured around different training sessions conducted by CNM staff and other experts, and presentations by various political actors, including regional actors such as the regional associations of municipalities, and public and private municipal suppliers, as well as national ones such as CNM itself, federal government officials, congress representatives, and sometimes the President of Brazil. In addition to attending the presentations, local policymakers use the meetings to network with each other and with state and federal officials. Each national conference brings around 4,000 municipal representatives and 2,000 mayors, while the regional conferences attract around 200 local political leaders, of which approximately 50 are mayors. Thus, our experiments take place in a quite natural setting, where policymakers are used to receiving useful information.

C. *Identifying Target Policies*

All information we provided to policymakers in the experiments satisfied two main conditions. First, the policies we focused on were directly within the control, familiarity, and broadly stated interest of municipal officials. Second, the information we provided was based on rigorous research, with emphasis on studies that evaluated interventions in Latin American countries.

To identify policy areas of interest to local policymakers, we conducted comprehensive surveys and focus groups with 60 mayors in May 2016. Substantial interest in acquiring research information was reported by mayors, especially on preschool education, preventive health care, and management practices. Mayors were also concerned with budgetary issues, especially considering the fiscal crisis affecting state and local governments in Brazil at the time (Mulas-Granados 2017). Based on mayors' priorities, we searched for, and systematically reviewed, research studies on Google Scholar, and the websites of J-PAL, IPA, 3ie, World Bank, IADB, and leading policy and research institutions in Brazil such as the repository of papers on IPEA, C-Micro-FGV, and on the websites of leading Brazilian scholars.

[13] The 12 regional conferences were held in the following states: Alagoas, Bahia, Ceará, Espírito Santo, Maranhão, Mato Grosso do Sul, Minas Gerais, Paraná, Piauí, Rio Grande do Sul, Santa Catarina, and São Paulo.

We identified a number of promising options, and after consulting with CNM, we decided to build the experimental interventions based on research information on early childhood development programs and on tax reminder letters. These policies were appealing for our purposes because they were evaluated in existing, rigorous research, and the taxpayer reminder letter policy we focus on in the policy-adoption experiment is inexpensive and relatively easy to implement. In addition, the set of studies evaluating the impact of each of the two policies varied in their attributes, allowing us to investigate how study features such as sample size and location affect policymakers' responses. We chose two distinct policies for the two experiments, since the beliefs experiment was largely conducted in between the intervention and the endline survey of the policy-adoption experiment. Since we did not want to contaminate the policy-adoption experiment, we were forced to choose a new policy topic when the opportunity to conduct the beliefs experiment arose.

How might the policies and research information we provide relate to mayors' objectives and constraints? The policy tool whose surprising effectiveness we describe in the policy-adoption experiment, tax reminder letters, has the potential to increase municipal revenues, easing the budget constraint that the mayor faces. Reminder letters are themselves a quite inexpensive tool, with a relatively low opportunity cost in terms of municipal resources. There are also good reasons to think that mayors would care about the effectiveness of ECD programs. In addition to any prosocial motivations, there is evidence from Brazil that voters reward or punish mayors based on their performance. For example, voters are less likely to reelect mayors who failed to improve test scores in municipal schools (Firpo, Pieri. and Souza 2017), or those who were exposed as being corrupt (Ferraz and Finan 2008). Mayors also appear to engage in competition with their neighboring municipalities on school performance (Terra and Mattos 2017). Given that mayors have a limited budget, it seems reasonable that information on effectiveness (and therefore cost-effectiveness) could be valuable to mayors.

II. Beliefs Experiment

In this section, we describe an experiment to measure (i) whether Brazilian policymakers demand research information, and (ii) how receiving such information affects their beliefs. The policy area this experiment focused on was Early Childhood Development (ECD) programs, a well-studied topic in social science. We find that policymakers value research on the effect of ECD programs, and update their beliefs substantially in response.

A. *Experimental Setting and Sample*

We implemented the beliefs experiment with 764 officials from 579 municipalities at 14 CNM meetings across Brazil in 2017 and 2018.[14] The conferences were

[14] The meetings comprised two national conferences held in Brasília (May 2017 and 2018), and12 regional *Diálogo Municipalista* conferences organized from August to December 2017 in the states of Alagoas, Bahia, Ceará, Espírito Santo, Maranhão, Mato Grosso do Sul, Minas Gerais, Paraná, Piauí, Rio Grande do Sul, Santa Catarina, and São Paulo. In addition, another group of 134 municipal officials from 117 municipalities also completed a survey on the advantages and disadvantages of the different studies used in this experiment.

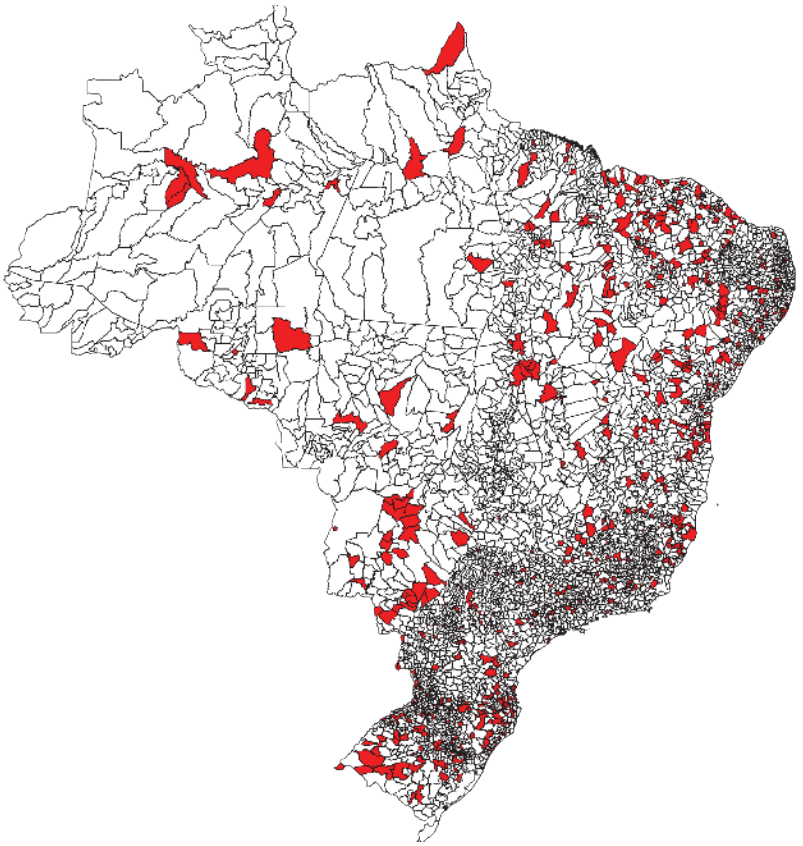

FIGURE 1. BELIEFS EXPERIMENT: SAMPLE MUNICIPALITIES

attended by mayors, vice-mayors, municipal secretaries, and local legislators. We designed a half-hour-long experiment that was privately self-administered by participants using tablets. The experiment was not announced in advance to participants. Instead, research assistants recruited conference participants during breaks in between sessions, as described in the next section. One of the researchers and one research assistant were present throughout to monitor and answer questions.

Almost 50 percent of participants in the experiment were mayors; 26 percent were local legislators; 18 percent were municipal secretaries; and 6 percent were vice-mayors. The geographical distribution of the municipalities represented is shown in Figure 1, and Table 1 displays summary characteristics.[15] About 38 percent of represented municipalities have mayors affiliated to a leftist political party, and approximately 20 (78) percent of children aged 0 to 3 (4 to 5) years old in these municipalities attend a preschool educational establishment. A total of 42 percent of participants report that their municipalities have implemented ECD programs.

[15] Demographic data are available from the Brazilian Statistical Office (IBGE 2016; AtlasBR 2010). Brazil's Superior Electoral Court provides data on electoral outcomes and mayors' characteristics (TSE 2012). Data on the education of the public administration were obtained from RAIS (Relação Anual de Informações Sociais), which is collected and annually compiled by the Brazilian Ministry of Labor (RAIS 2012). Data from Base Dos Dados enabled the linking across databases (Carabetta et al. 2020).

Table 1—Beliefs Experiment: Summary Stats and Balance

Variables	Mean control	Δ Developing	*p*-value	Δ Large	*p*-value
Mayors' characteristics					
Male	91.46	−1.32	0.38	−4.04	0.00
Age	48.61	−0.11	0.82	−0.49	0.28
College	57.62	2.80	0.20	0.13	0.95
Second term	18.29	0.44	0.82	−0.07	0.97
Electoral margin victory	14.19	0.18	0.81	0.36	0.55
Leftist political party	38.72	−0.99	0.64	1.80	0.42
Municipalities' characteristics					
Population	24.49	1.45	0.48	1.24	0.40
College population	4.915	−0.07	0.52	0.02	0.87
Public adm. college	34.16	−0.97	0.09	−0.85	0.17
Poverty	26.45	0.48	0.55	0.41	0.61
Gini	49.48	0.48	0.09	0.44	0.13
Big South	51.22	−0.75	0.74	−1.07	0.63
Per capita income	457.1	−8.79	0.40	1.02	0.93
Kids in school (0–3)	19.88	−1.04	0.04	0.08	0.87
Kids in school (4–5)	78.34	−0.41	0.54	0.16	0.83
ECD policy survey characteristics					
Mayor	49.70	−0.16	0.94	−1.33	0.55
Prof. politician	29.27	0.74	0.72	−0.47	0.81
Leftist scale	23.78	−2.37	0.19	−1.79	0.35
Implemented ECD	41.77	0.40	0.85	−3.02	0.17
Heard ECD	26.22	−0.81	0.68	−0.11	0.95

Notes: Sample means of control observations. Δ *Developing* and Δ *Large* report the estimated coefficient, with its respective *p*-value, of a linear regression of each characteristic of the mayor, the municipality and the ECD policy survey, on two dummy variables: a dummy equal to 1 for Jamaica and Colombia and 0 otherwise (*Developing*), and on a dummy equal to 1 for Colombia and United States and 0 otherwise (*Large*). Control observations are those for which the dummy *Developing* and the dummy *Large* are equal to 0. The linear regression is estimated with 1,368 observations. Robust standard errors are clustered at the individual level (764 clusters). The mean of control is calculated with 368 observations. The first block of variables reports characteristics of the mayor who runs the municipality. *Leftist political party* (=1 for mayors belonging to a center-leftist party according to historical political platforms, 0 otherwise). The second block of variables reports characteristics of the municipality. *Population* is the municipality's number of inhabitants (in thousands). *College population* is the municipality's share of adults with college degrees. *Public administration college* is the share of municipal public employees with college degrees. *Poverty* is the municipality's poverty rate. *Gini* is the municipality's Gini coefficient. *Big South* is equal to 1 for municipalities in the south, southeast and midwest regions, 0 otherwise. *Per capita income* is the municipality's monthly income per capita. *Kids in school* (0–3) is the share of kids 0–3 years old in the municipality who attend preschool education. *Kids in school* (4–5) is the share of kids 4–5 years old in the municipality who attend preschool education. The third block of variables reports characteristics self-reported by participants in the survey experiment. *Professional politician* is equal to 1 if the participant occupied an elective position in the previous term, 0 otherwise. *Leftist scale* is equal to 1 if the participant self-identified as leftist (0–4) on a 0–10 scale, 0 otherwise. *Implemented ECD* is equal to 1 if the participant reported that the municipality implemented a ECD program before, 0 otherwise. *Heard ECD* is equal to 1 if the participant reported that he/she had heard about ECD programs before, 0 otherwise.

We recruited 38 percent of attending mayors, 49 percent of vice-mayors, 35 percent of municipal secretaries, and 41 percent of local legislators. Participation was limited by the number of tablets available and the limited breaks in the conference schedule, but participants may also have selected into the study based on their interest in the participation incentive (lottery tickets), or their interest in education policy. The latter would potentially bias our estimates of demand upward. Online Appendix Table A.1 shows that participating mayors were 7 percentage points more likely to be from leftist parties than nonparticipants, but were otherwise similar on a range of other characteristics.

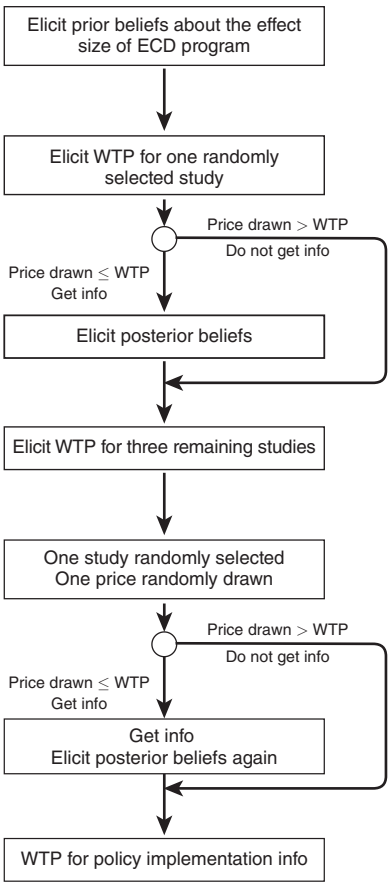

FIGURE 2. BELIEFS EXPERIMENT: STRUCTURE

B. *Experimental Design*

The structure of the experiment, depicted graphically in Figure 2, was as follows. We began by introducing the ECD policy. Then we elicited participants' prior beliefs about the effectiveness of the policy, and their willingness-to-pay (WTP) to learn the findings from related impact evaluation research. Next, we revealed the findings, and finally, we elicited participants' posteriors to assess the extent to which the research findings affected their beliefs. Online Appendix Section B provides the key parts of the experimental script. Below, we describe the experiment in greater detail.

Introductory Stage.—We began with a short survey eliciting demographic and professional information. Next, we described ECD programs, highlighting the key outcomes on which such programs are evaluated (test scores, cognitive skills) and how those outcomes are reported (standardized effect sizes). To ease understanding of the policy and its objectives, we provided illustrative examples of current similar programs in Brazil and presented participants with a few benchmarks for effect sizes, such as the gains in standardized test scores associated with an additional year of high school in Brazil (0.2 SD).

Eliciting Priors.—We began the main part of the experiment by eliciting the participant's prior beliefs. Specifically, we asked what they believed the impact of the policy on cognitive skills was likely to be if it were to be implemented in his/her own municipality.[16] Immediately after, we asked a similar question about the expected impact in two other locations. These two other locations were randomly chosen out of four locations where academics have estimated the impact of ECD programs using RCTs. These studies vary in location and sample size. They evaluate comparable ECD programs in Colombia ($n = 1{,}420$) (Attanasio et al. 2014), Jamaica ($n = 130$) (Grantham-McGregor et al. 1991; Walker et al. 2005), Michigan ($n = 123$) (Schweinhart et al. 2005; Barnett 2011), and across multiple states in the United States ($n = 4{,}667$) (Puma et al. 2010). When the relevant studies were presented to the participant, we highlighted both the study location and sample size.[17]

Attributes	Small sample	Large sample
Developing country	Jamaica, $n = 130$	Colombia, $n = 1{,}420$
Rich country	Michigan, $n = 123$	United States, $n = 4{,}667$

While we cannot incentivize accurate beliefs about the impact in the participant's own municipality (since we do not observe the true effect), we randomize incentives to accurately predict the effect in the other two locations (where we can compare the participant's prediction to the estimates from the research). In practice, we found that the size of the incentives has no effect on priors, WTP, or posteriors, suggesting that participants took the questions seriously even in the absence of incentives, and that making better predictions for the sake of higher payoff *within the experiment* is not an important driver of this paper's results.

WTP and Belief Updating: Round 1.—After the participants reported their priors, we offered them the chance to purchase the findings (i.e., learn the estimated effect size) from one randomly chosen study. The experimental currency in which we elicited WTP consisted of lottery tickets, which also incentivized participation. We initially endowed each participant with 100 such lottery tickets, each with a chance of winning a free trip to visit the United States (typically a visit to Boston, including a tour of the Harvard University campus). Participants could save their lottery tickets for the lucky draw or use some, or all of them, to learn the estimated effect size of the study. Following a Becker-DeGroot-Marschak elicitation procedure (BDM), we measured the participant's maximum WTP [0 to 100] to find out the results of the relevant study. We then drew a randomized price for the study. If the price was below the participant's WTP, we revealed the findings and deducted the price from the participant's stock of lottery tickets.

[16] For simplicity and due to limited time with each participant, we elicited only point predictions (about effects on cognitive skills), rather than full probabilistic beliefs. This is an important limitation of the study, which we return to later.

[17] We did not use the labels *Developing country* or *Rich country*, nor *Small sample* or *Large sample*. We simply presented the location and the sample size.

To deal with the issue of selection into seeing a study result based on one's WTP, while maintaining incentive-compatibility in the BDM procedure, the price was drawn from a distribution with high mass at zero. Consequently, 80–90 percent (depending on the conference) of participants received the information *regardless of their WTP*. Whenever presenting results on belief updating, we also present results for this subsample, which receives the information without selection. This approach also has the advantage that we get to observe belief-updating for most participants.

For those who received the information, we subsequently elicited posterior beliefs about the expected impact of the policy in their own municipality, and in a study location that was *not* offered for purchase in this round. We do not ask for an updated posterior from participants who do not receive a study's results. As is standard in lab experiments on belief updating, we assume that beliefs do not change over the matter of minutes in the absence of new information (e.g., see the papers reviewed in Benjamin 2019 or Vivalt and Coville 2020; Mobius et al. 2011; Eil and Rao 2011).[18]

WTP and Belief Updating: Round 2.—In the next stage, we presented the participant with a menu of the three studies that were not offered for purchase in round 1, again highlighting each study's location and sample size. The participant received a fresh budget of 100 lottery tickets and was told that one of the three studies would be randomly offered for purchase. They were asked to report their WTP for *each* study, to be implemented if that study was randomly chosen for sale. We thus obtained incentive-compatible WTPs for each of the three studies. We revealed the findings of one study following the same procedure as before, and again elicited an updated posterior belief. Having this second round allows us to observe a second instance of belief-updating per participant, increasing statistical power. It also allows us to learn how the weight placed on research information diminishes from the first to the second study on the topic.

C. *Results*

We interpret the results through the lens of a simplified Bayesian-learning framework. Suppose that policymaker i has a prior belief $S_i^{pr} \sim \mathcal{N}\left(\mu_i^{pr}, \Sigma_i^{pr}\right)$, where μ_i^{pr} is the mean of i's prior and Σ_i^{pr} is the perceived variance or uncertainty of their prior about the likely effect of the ECD policy if implemented in their municipality. The effect size from the research study can be thought of as a noisy signal $S_{i,c}^{I} \sim \mathcal{N}\left(\mu^{I}, \Sigma_{i,c}^{I}\right)$, drawn from a distribution centered around the true value μ^I, but with variance $\Sigma_{i,c}^{I}$, where c indexes characteristics of the study, such as its sample size or location. Then, a Bayesian policymaker who wants to have accurate beliefs (to minimize mean squared error) will form a posterior S_i^{po}:

$$S_i^{po} = \left(1 - \pi\right) S_i^{pr} + \pi S_{i,c}^{I},$$

[18] This assumes away the possibility that simply being asked a second time would cause a systematic shift in beliefs, for instance due to thinking harder. Under our assumption, the change in belief from prior to posterior is the treatment effect of learning the information.

with the weights $\pi = \Sigma_i^{pr} / \left(\Sigma_i^{pr} + \Sigma_{i,c}^{I}\right)$. That is, a Bayesian learner's posterior will be a convex combination of their prior and the "signal" (i.e., the effect-size from the study), with weights proportional to the perceived relative precision of each component. While we cannot test the assumptions of this model, particularly the normally distributed probabilistic beliefs, since we only measure point beliefs, this framework provides a useful benchmark for the belief-updating we study.

We can think of the key attributes of the study, location and sample size, as affecting the perceived precision or informativeness of the noisy signal. If participants think that larger-sample studies are more informative ($\Sigma_{i,large}^{I} < \Sigma_{i,small}^{I}$), they will place greater weight on the effect size of larger-sample studies while forming their posterior beliefs. Importantly, if policymakers value having accurate beliefs about the effectiveness of ECD policies, their WTP for signals will be higher for the signals which they will ex post weight more strongly in their belief updating.

Priors about Effect Size.—We start by analyzing policymakers' priors about the effectiveness of ECD policies. The average policymaker prior appears sensible, if a bit optimistic. Online Appendix Table A.2 shows that the average policymaker believes that ECD policies are more effective in rich countries (effect size of 0.45–0.50 SD) than in developing countries (effect size of 0.37–0.42 SD). On average, municipal officials believe the effect size in their own municipality (0.42 standard deviations) is very close to the average prior for the developing countries. However, this masks substantial heterogeneity in priors: the standard deviation of priors is 0.22, implying substantial disagreement across policymakers.[19] Since we only elicit point beliefs rather than probabilistic beliefs, we do not have a measure of the uncertainty in each policymaker's beliefs.

Willingness-to-Pay for Estimated Effect Size.—After policymakers reported their priors, we elicited their WTP to learn the research finding of one of the four (randomly assigned) studies. If policymakers value accurate beliefs, WTP should be larger the more informative the signal is perceived to be. We estimate the following equation:

$$WTP_{ijs} = \beta_0 + \beta_1 Developing_{ijs} + \beta_2 Large_{ijs} + \varepsilon_{ijs}, \tag{1}$$

where WTP_{ijs} is the WTP (in terms of lottery tickets) for the research finding of policymaker i in round $j \in 1,2$ for study $s \in$ Michigan, United States, Jamaica, Colombia. The variable $Developing_s$ equals 1 for studies in Jamaica or Colombia and 0 otherwise; $Large_s$ equals 1 for the two large-sample studies (Colombia with $n = 1{,}420$ and United States with $n = 4{,}667$) and 0 otherwise (Jamaica with $n = 130$ and Michigan with $n = 123$). Standard errors are clustered at the individual level.

Table 2 presents the OLS results from specification (1). Column 1 pools the two rounds, while columns 2 and 3 present estimates separately for round 1 and round 2 respectively. We find that policymakers allocate on average 45 lottery tickets (out of

[19] Of course, some of this variance in priors may reflect noise in the belief-elicitation process.

Table 2—Beliefs Experiment: Willingness to Pay by Study Characteristics

	WTP		
LHS variable	(1)	(2)	(3)
Large	3.8221	2.3554	4.4182
	(0.7912)	(2.3944)	(1.0152)
Developing	0.3783	1.5948	−0.2735
	(0.7907)	(2.3951)	(1.0039)
Observations	2,573	764	1,809
Round	1 and 2	1	2
Clusters (individuals)	764	764	604
Mean LHS	44.62	48.39	43.03
SD LHS	31.77	33.06	31.09

Notes: OLS results. The dependent variables are willingness to pay, which are elicited in two different rounds. *Developing* is a dummy which is equal to 1 for Jamaica and Colombia, 0 otherwise. *Large* is a dummy which is equal to 1 for Colombia and United States, 0 otherwise. Difference in number of clusters between columns 2 and 3 is due in part to a different experimental design of last CNM conference, in which only one study was offered for purchase. *Mean LHS* is the mean WTP on the left-hand side of each equation. *SD LHS* is the standard deviation of WTP on the left-hand side of each equation. Robust standard errors clustered at the individual level are in parentheses. *p*-value of *Large* (column 2) = *Large* (column 3) test is 0.484. *p*-value of *Developing* (column 2) = *Developing* (column 3) test is 0.524.

the 100 tickets they are endowed with each round) to learn about the effect size of a particular study. While this is a large share of their experimental endowment, it is difficult to interpret the level directly since the currency is lottery tickets, whose subjective value is unobserved. To benchmark the WTP, we calculated a money metric for the experimental currency by offering gift cards from a major retail and online chain (*Lojas Americanas*, similar to Walmart) for purchase using a similar BDM procedure to a subsample of participants. We found that an additional lottery ticket was exchanged for approximately US$0.80 worth of gift cards. This benchmarking must be interpreted with caution, but suggests that the baseline WTP for the research finding of 45 lottery tickets was equivalent to US$36, between 0.4 percent and 0.9 percent of a mayor's monthly wage. There is substantial heterogeneity in demand: the standard deviation of WTP is 32 lottery tickets. Yet, 99 percent of participants have strictly positive WTP.[20] WTP declines from round 1 to round 2: the second study a policymaker is offered is valued 11 percent less than the first.

We next analyze whether demand for research findings varies with the attributes of the research. We find that political leaders are willing to pay about 9 percent more for large-sample size studies than for smaller-sample studies. Thus, policymakers appear to ex ante value the statistical precision of a study. This relationship is stronger in the second round, when studies are offered side-by-side, but the second-round estimate is not statistically different from the first-round estimate (p-value 0.484). In

[20] Readers might wonder why participants would not simply look up the research themselves. While this may happen to some extent, we believe that unfamiliarity with research-information sources, language barriers, and difficulty interpreting academic writing are all factors that make this strategy difficult for our study participants. Our estimates may be thought of as capturing their WTP for simplified, conveniently presented, bottom-line information.

contrast, and contrary to our priors, we do *not* find significant differences between the WTP for research findings from Colombia and Jamaica versus Michigan or across the United States. This suggests that, on average, Brazilian policymakers do not consider studies from other developing countries to be more informative, more externally valid for them, than rich-country studies.

We report participant and municipality-level correlates of WTP in online Appendix Table A.3. Only 3 characteristics out of 20 are significantly associated with WTP in this exploratory analysis: whether the participant is male, whether their municipality has previously implemented an ECD policy, and whether they reported having previously heard about such policies despite not having implemented them. Through the lens of the framework, the latter correlations are not inevitable: policymakers with more past experience with a policy might have a more precise prior, and therefore not value additional information. Instead, we find that it is precisely the policymakers who implement and spend municipal resources on ECD programs who have the highest WTP for related research information. Presumably, this is because having accurate beliefs about such programs is more valuable to them. Term-limited mayors and those with a higher margin of electoral victory (who presumably face less electoral competition), in contrast, do not have higher WTP for research information.

Belief Updating.—Having established that political leaders value research findings, and pay more for larger-sample studies, we turn to whether and how they actually update their beliefs upon learning research findings. Note that if policymakers purchase information purely to use it to persuade others, for instance, they might not update their own beliefs upon receiving the information.

Following the Bayesian framework, we estimate the following equation:

$$Posterior_{ijs} = \beta_1 Prior_{ij} + \beta_2 Signal_{ijs} + \varepsilon_{ij}, \tag{2}$$

where $Posterior_{ijs}$ is policymaker i's updated belief about the likely effect in their own municipality after learning the effect size from study $Signal_{ijs}$ of study s in round j. Posteriors after round 1 serve as priors for round 2, and standard errors are clustered at the individual level.

Table 3 presents the OLS estimates of specification (2). Column 1 pools the two rounds, while columns 2 and 3 present estimates separately for round 1 and round 2, respectively. Consistent with the framework on average, $\hat{\beta}_1$ and $\hat{\beta}_2$ are both positive and statistically significant, and sum up to approximately 1. Participants place about two-thirds weight on their prior and one-third on the study finding on average, and do not simply accept or repeat back the research finding. This finding perhaps reduces concerns about experimenter demand effects. They place similar weight on the study finding when forming beliefs about their own municipality, compared to beliefs about an alternative location (column 4 versus column 2). They place more weight on their prior in the second round, when it already incorporates the finding of the first study they received. Put differently, the weight placed on a study's findings falls by 30 percent from the first to the second study a policymaker learns about. As described previously, by design, 80–90 percent of participants are assigned a zero price and receive the research information regardless of their WTP. Column 5

Table 3—Beliefs Experiment: Belief Updating

	Posterior				
LHS variable	(1)	(2)	(3)	(4)	(5)
Prior	0.6824	0.5902	0.7902	0.6528	0.6813
	(0.0214)	(0.0295)	(0.0237)	(0.0280)	(0.0224)
Signal	0.3230	0.3749	0.2607	0.3622	0.3209
	(0.0194)	(0.0261)	(0.0234)	(0.0293)	(0.0203)
Observations	1,240	700	540	543	1,131
Round	1 and 2	1	2	1	1 and 2
Beliefs about	Municipality	Municipality	Municipality	Random study	Municipality
Received study for free	No	No	No	No	Yes
Clusters (individuals)	755	700	540	543	731

Notes: OLS results. The dependent variables are posterior beliefs, which are declared after successfully buying the results from a study in each round. *Prior* is the belief of the respondent about the effect, right before buying a study. *Signal* is the bought study's effect size. When dealing with a second update in posteriors, the first update is treated as a prior. In the rows below the coefficients, *Beliefs about* specifies which location the beliefs are elicited for, either the respondent's own municipality (columns 1, 2, 3, and 5) or one of the four possible study locations (column 4). *Received study for free* indicates whether participant received the information regardless of their WTP. Difference in clusters between columns 2, 3, and 4 is due in part to a different experimental design of last CNM conference, in which only one study was offered for purchase. Robust standard errors clustered at the individual level are in parentheses.

restricts attention to these observations, and finds very similar results as in the full sample.[21]

Under the assumption that beliefs do not change absent any new information, we can also report the effect of receiving each study. In round 1, the average posteriors after receiving each study are 0.49 (Michigan), 0.51 (Jamaica), 0.37 (Colombia), and 0.35 (United States). Compared to the average prior in each case, this implies treatment effects of $+0.075$ for Michigan, $+0.096$ for Jamaica, -0.034 for Colombia, and -0.073 for the large-sample US study respectively.

Online Appendix Tables A.5 and A.6 report an exploratory analysis of heterogeneity in belief updating by mayors' and municipalities' characteristics. College-educated and leftist mayors place less weight on their priors and more weight on the study finding. Older mayors do the reverse: they update their beliefs less when faced with research information. While mayors who have implemented ECD programs had higher WTP for studies, as described above, they do not update more based on them. Finally, reelection incentives and political competition do not have a systematic relationship with updating. Just as term-limited mayors and those with larger electoral margins of victory did not have lower WTP, they also do not place lower weight on the research findings. Of course, these findings cannot be interpreted causally, and they should be treated as suggestive at best.

[21] One concern is that the prior may be measured with noise, and that such measurement error will attenuate the coefficient on Prior. We can address this issue by instrumenting for the prior in round 2 using the revealed study in round 1. Online Appendix Table A.4 contrasts the weights on the priors in updating using the OLS specification (column 4) and a 2SLS specification where the prior is instrumented (column 5). The coefficients on Prior are very similar, suggesting that the attenuation bias problem is not severe in practice.

In order to test whether participants update more based on large-sample or developing-country studies, we estimate

$$
\begin{aligned}
(3)\quad Posterior_{ijs} = {} & \beta_1 Prior_{ij} + \beta_2 Signal_{ij} \\
& + \beta_3 Developing_{ijs} \times Prior_{ij} + \beta_4 Developing_{ijs} \times Signal_{ij} \\
& + \beta_5 Large_{ijs} \times Prior_{ij} + \beta_6 Large_{ijs} \times Signal_{ij} + \varepsilon_{ij},
\end{aligned}
$$

where $Large_{ijs}$ and $Developing_{ijs}$ are defined as in equation (1). Under the framework, if an individual perceives a study to be more informative, they will place more weight on the signal from that study and correspondingly less weight on their prior. Therefore, to test whether participants perceive (say) large-sample studies to be more informative, we can test whether $\beta_5 < 0$ and $\beta_6 > 0$, or instead (a weaker test) whether $\beta_6 - \beta_5 > 0$.

Table 4 presents the OLS results of specification (3). Again, column 1 pools the two rounds, while columns 2 and 3 present estimates separately for each round. We find consistent evidence that participants place greater weight on signals from large-sample studies, but not on signals from developing-country studies. This lines up with the findings on WTP, and confirms that these policymakers find larger-sample studies to be more informative, but do not consider studies from developing and rich countries to be differentially informative. The greater weight placed on large-sample studies is evident also in round 1, when one study is presented in isolation. The pattern of results holds up, and indeed is slightly strengthened, when we restrict attention to cases where the price drawn was zero in column 5.

Figure 3 depicts the observed belief updating using binned scatter plots.[22] The y-axis plots the size and direction of updating $(Posterior - Prior)$ for a given news shock due to the signal $(Signal - Prior)$ on the x-axis. Panel A includes all instances of updating, pooling across studies and rounds, and adds a linear OLS fit. A few points are noteworthy. First, the relationship does appear to be linear, in line with the Bayesian model and our empirical specification in (2). Second, there is no evidence of asymmetric (optimistic) updating, which would show up as a kink at the origin with a steeper slope to the right of zero. The other panels in turn depict updating separately for large- and small-sample studies (panel B) and for rich- and developing-country studies (panel C). The stronger updating response to large-sample studies is evident, as is the similar average response to rich and developing-country studies.

Figure 4 plots the histogram of the belief-updating responses. Specifically, for each instance of updating, we calculate $\pi = (Posterior - Prior)/(Signal - Prior)$ and then average these responses within each individual. The figure reveals substantial heterogeneity in the weight placed on the research findings. Twenty-eight percent of policymakers appear to ignore the study result and do not update their beliefs at all ($\pi = 0$). Forty-three percent of policymakers have updating weights strictly between 0 and 1. Fifteen percent update in the wrong direction ($\pi < 0$)

[22] Online Appendix Figure A.1 presents the corresponding figures separately by study.

Table 4—Beliefs Experiment: Belief Updating: Weight Placed on Large-Sample and Developing-Country Studies

	Posterior				
LHS variable	(1)	(2)	(3)	(4)	(5)
Prior	0.6388	0.5600	0.7509	0.6685	0.6420
	(0.0368)	(0.0531)	(0.0471)	(0.0543)	(0.0384)
Signal	0.3306	0.3780	0.2653	0.3351	0.3280
	(0.0284)	(0.0397)	(0.0384)	(0.0429)	(0.0299)
Prior × developing	−0.0093	−0.0247	−0.0106	−0.0920	−0.0083
	(0.0389)	(0.0599)	(0.0477)	(0.0574)	(0.0414)
Signal × developing	0.0091	0.0082	0.0189	0.0682	0.0039
	(0.0349)	(0.0515)	(0.0472)	(0.0578)	(0.0367)
Prior × large	−0.0535	−0.0904	−0.0307	−0.0563	−0.0663
	(0.0480)	(0.0690)	(0.0600)	(0.0714)	(0.0501)
Signal × large	0.3233	0.4068	0.2413	0.2744	0.3510
	(0.0712)	(0.0963)	(0.0942)	(0.1176)	(0.0745)
Observations	1,240	700	540	543	1,131
Round	1 and 2	1	2	1	1 and 2
Belief about	Municipality	Municipality	Municipality	Random study	Municipality
Received study for free	No	No	No	No	Yes
Clusters (individuals)	755	700	540	543	731
p-value prior × developing = signal × developing	0.791	0.755	0.742	0.142	0.869
p-value prior × large = signal × large	0.001	0.002	0.064	0.069	0.001

Notes: OLS results. The dependent variables are posterior beliefs, which are declared after successfully buying the results from a study in each round. *Prior* is the belief of the respondent about the effect, right before buying a study. *Signal* is the bought study's effect size. When dealing with a second update in posteriors, the first update is treated as a prior. *Developing* is a dummy which is equal to 1 for Jamaica and Colombia, 0 otherwise. *Large* is a dummy which is equal to 1 for Colombia and United States, 0 otherwise. In the rows below the coefficients, *Beliefs about* specifies which location the beliefs are elicited for, either the respondent's own municipality (columns 1, 2, 3, and 5) or one of the four possible study locations (column 4). *Received study for free* indicates whether participant received the information regardless of their WTP. Difference in clusters between columns 2, 3, and 4 is due in part to a different experimental design of last CNM conference, in which only one study was offered for purchase. Robust standard errors clustered at the individual level are in parentheses.

while 13 percent overreact ($\pi > 1$). This distribution appears quite similar to that found in Vivalt and Coville (2020), who present participants at a World Bank impact evaluation workshop with a hypothetical study in a belief-updating exercise. They also find a substantial share of participants who do not update ($\pi = 0$), and about 55 percent of participants displaying ($0 \leq \pi \leq 1$). The average updating weight in our sample, about 0.37 in the first round, is also comparable to the median weight of 0.5 found by Vivalt and Coville (2020).

What explains the approximately one-quarter of participants who do not respond to the information? One possibility is simply inattention or effort-minimization by participants. However, attention checks ensured that participants at least briefly registered the study findings, and participants were required to actively report a posterior during each belief elicitation. The interpretation through the lens of the model would instead be that these policymakers have very confident priors, and therefore think that research is uninformative. This is possible, but is at least somewhat inconsistent with 99 percent of participants having a positive WTP for study results.

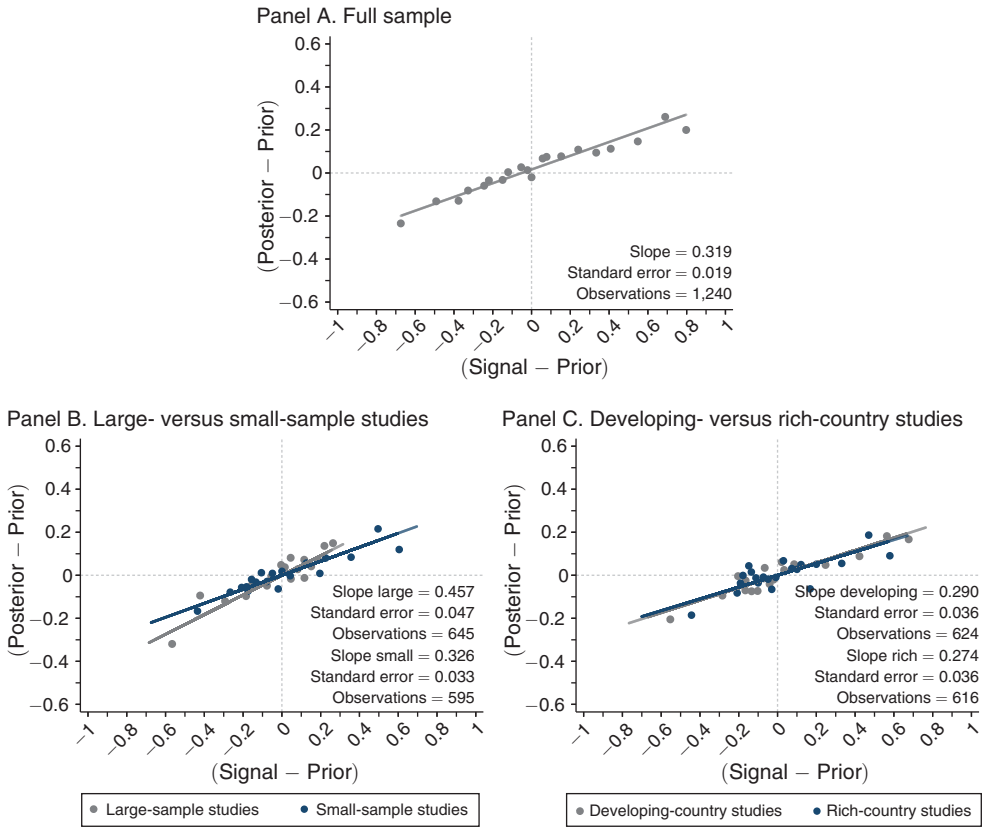

FIGURE 3. BELIEFS EXPERIMENT: BELIEF UPDATING

Notes: Comparison between the difference in respondent's perceptions after buying a study (i.e., posterior beliefs minus prior beliefs) and the difference in respondent's perceptions before buying a study (i.e., signal minus prior beliefs), averaged over bins of rounds 1 and 2. *Prior* is the belief of the respondent about the effect, right before buying a study. *Signal* is the bought study's effect size. When dealing with a second update in posteriors, the first update is treated as a prior. Panel A shows statistics for full sample. Panels B and C compare statistics between large- versus small-sample studies (controlling for the location attribute of the study) and between developing- versus rich-country studies (controlling for the sample size attribute of the study), respectively. *Large-* (*Small-*)*sample studies* include Colombia and United States (Jamaica and Michigan); while *Developing-* (*Rich-*)*country studies* include Colombia and Jamaica (Michigan and United States). The slope and robust standard errors clustered at the individual level are based on a linear regression with a constant term.

Another possible factor is rounding issues in belief measurement in our experiment. Beliefs could only be reported at intervals of 0.1 SD. Thus, underlying belief updates from, for example, 0.46 to 0.54 will appear to involve no update at all, if both are rounded to 0.5. These rounding issues can also inflate the share who appear to overreact, since updating from 0.44 to 0.46 may be measured as an update of 0.1. This is a major caveat to the interpretation of the individual-level updating distribution.[23]

[23] There is at least one other factor that may explain updating in the wrong direction or overreaction. Participants may consider different studies to not just be more or less noisy signals, but also to have some known bias, e.g., thinking that programs in the US are better implemented, such that one should subtract some fixed number from the effect size, or conversely thinking that effects in developing countries will be larger since they are further away from

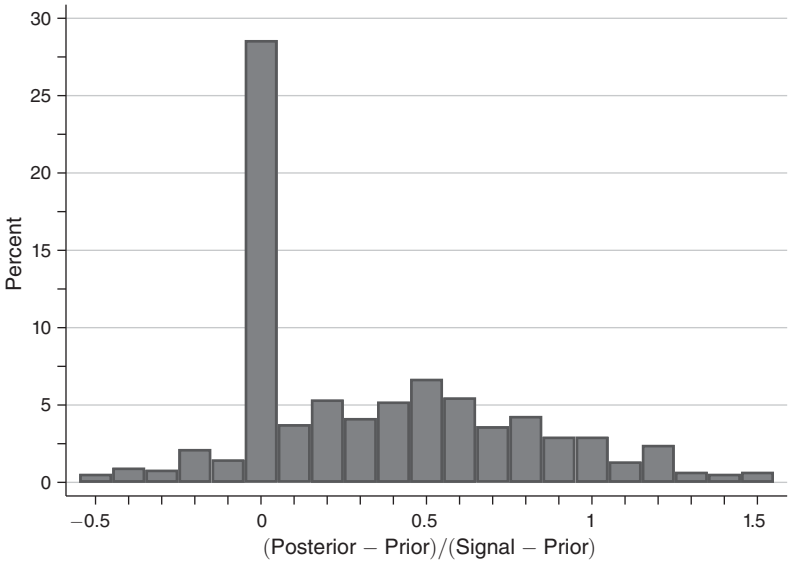

FIGURE 4. BELIEFS EXPERIMENT: BELIEF UPDATING DISTRIBUTION

Note: Distribution of the share of the difference in respondent's perceptions after buying a study (i.e., posterior beliefs minus prior beliefs) and the difference in respondent's perceptions before buying a study (i.e., signal minus prior beliefs), averaged within respondents' rounds 1 and 2.

This concern is likely to matter less when measuring average responses across many participants, which most of our analysis focuses on. Nonetheless, online Appendix Table A.4 provides the belief-updating regressions while consecutively dropping participants who never update ($\pi = 0$), or excluding those with $\pi \leq 0$ or $\pi \geq 1$.

One natural question is whether the patterns of belief updating we observe are quantitatively sensible and in line with rational Bayesian updating on average. Since we do not measure the precision of beliefs or provide participants with the precision of the signal, we cannot calculate how much a Bayesian *should* update. Therefore, we cannot say with confidence whether the extent of updating is about right, or instead too much or too little on average. Nor do we judge whether it is appropriate for policymakers to place equal weight on results from rich and developing countries. Our results along these lines are purely descriptions of how policymakers *do* update.

However, with additional assumptions, it is possible to shed some light on a related question: is the response to sample size in line with a Bayesian model? Let us first consider the updating weights placed on the different kinds of studies. Suppose that priors, signals, and posteriors are all normal, and that the precision of signals depends only on sample size. Then one can show that, for a Bayesian, the ratio of the updating weights placed on two studies *j* and *k* should be

the efficient frontier. Then, seeing a US result slightly greater than one's prior regarding one's own municipality could actually cause one to update in the opposite direction.

closely related to their sample sizes: $\Pi_j/\Pi_k = n_j/n_k$, where $\Pi_i \equiv \pi_i/(1-\pi_i)$ is the ratio of the optimal weights placed on the signal and the prior and n_i denotes the sample size of study i.[24] To compare this to the actual empirical weights, we can calculate the average weight placed on the signal for each study by estimating (2) separately by study (in round 1, and normalizing the weights on the prior and signal to sum to one). Comparing the two developing-country studies, we find that $\Pi_{Colombia}/\Pi_{Jamaica} = 1.8 < 10.9 = n_{Colombia}/n_{Jamaica}$. This implies that the weight placed on the larger-sample Colombia study relative to the smaller-sample Jamaica study is less than what Bayesian learning would justify. We find a similar qualitative pattern in the case of the two rich-country studies: $\Pi_{US}/\Pi_{Michigan} = 5.4 < 37.9 = n_{US}/n_{Michigan}$. In both cases, policymakers underreact to variation in sample sizes relative to a Bayesian. It is important to emphasize that we cannot say whether this is because too little weight is placed on the large-sample study, or too much on the small-sample study, or both.[25] Our finding is consistent with an extensive literature on sample-size neglect (Kahneman and Tversky 1972, Griffin and Tversky 1992, Benjamin 2019) and with Vivalt and Coville (2020), who find that policymakers underreact to the size of confidence intervals.

Next, consider the difference in WTP between the large and small-sample studies. Is this quantitatively justified by the subsequent differences in updating weights? Suppose WTP is proportional to the expected reduction in the policymaker's mean-squared prediction error. Then, one can show that $WTP_s \propto \pi_s \Sigma^{pr}$. For a Bayesian, we should therefore have $WTP_i/WTP_j = \pi_i/\pi_j$. Empirically, comparing the large-sample and small-sample studies (again normalizing the weights on prior and signal), we have $WTP_{large}/WTP_{small} = 1.09 < 1.53 = \pi_{large}/\pi_{small}$. Thus, while policymakers' WTP does respond to sample size, the sensitivity of WTP to sample size is lower than that justified by the sensitivity of belief updating (which itself may be too low, as described above). Again, we cannot say whether this is because the WTP for the large-sample studies is too small, or instead because the WTP for small-sample studies is too high. For instance, it could be that the baseline WTP is inflated because it captures not just concern for the informativeness of the study, but also some experimenter demand effects or just confusion.

Caveats, Confounds, and Qualifications.—While we interpret the differences in WTP and belief-updating across sample size and study location as the direct effect of these two characteristics, both could be correlated in policymakers' minds with omitted variables such as the quality of the research, the scale of implementation of the program, etc. To shed light on this, we conducted a debriefing survey with a

[24] The steps in the argument are as follows. The ratio of the variance of the signals coming from two studies j and k is $\Sigma_j^I/\Sigma_k^I = n_k/n_j$, where n denotes sample size. Recall that the optimal weight placed by a Bayesian on a signal from study j is $\pi_j = \Sigma^{pr}/(\Sigma^{pr} + \Sigma_j^I)$. Rearranging, we get that $\Sigma^{pr} = \Pi_j \cdot \Sigma_j^I$, where we have defined $\Pi_i \equiv \pi_i/(1-\pi_i)$. Since the priors are the same across studies due to randomization, it must be that $\Pi_j/\Pi_k = \Sigma_k^I/\Sigma_j^I$. Since $\Sigma_k^I/\Sigma_j^I = n_j/n_k$, it follows that $\Pi_j/\Pi_k = n_j/n_k$. The empirical analogues for this expression are restricted to updating in round 1, when randomization ensures that priors are the same on average.

[25] There is another way in which the observed updating clearly departs from the Bayesian learning framework: large-sample studies lead to an increase in weight placed on the signal but the reduction in weight placed on the prior is not equal and opposite but is instead smaller, as shown in Table 4.

subset of the sample ($n = 294$). We find that 59 percent of policymakers who preferred large-sample studies chose statistical precision as the reason. Intriguingly, a smaller share also reported preferring larger-sample studies because they are more likely to have evaluated programs implemented at scale (23 percent) and by the government (15 percent). In the case of study location, the survey results are more mixed: while individuals who preferred studies from Colombia or Jamaica reported their lower standard of living and similar state capacity as reasons, a substantial share also reported preferring the US studies, and listed a higher standard of living and similar state capacity as reasons. One interpretation is that some policymakers in Brazil may see their municipalities as closer to developing countries, while others may see themselves as closer to rich countries.

One glaring weakness is that we only consider studies from three countries. What we interpret as a "rich-country effect" could instead be a "USA effect"; Brazilian policymakers might not value research from other rich countries. Similarly, it could be that they would place much greater weight on evidence from Brazil, and consider it much more relevant than findings from Colombia or Jamaica. We were limited in our ability to explore these questions due to a lack of comparable studies from more countries, including Brazil.

The results on belief updating (but not WTP) have another potential confound in interpretation: the two larger-sample studies in practice estimated smaller effect sizes. This is a feature in the four studies we use, and also more generally documented in the ECD literature (Barnett 2011). What if participants simply update more (in proportional terms) in response to small effect sizes, say due to concerns about greater publication bias in small studies, or because large effect sizes seem implausible? We have some unplanned variation which may shed light on this concern: in 6 of the 14 conferences where the experiment was conducted, we reported a different (smaller) effect-size for certain studies. Specifically, for the small-sample studies alone, we reported the estimated effect sizes at a much longer time horizon (without flagging this discrepancy), which resulted in a smaller effect size. Online Appendix Table A.7 tests whether the larger weight on large-sample study signals is less pronounced in those conferences. Consistent with our initial interpretation, the weight placed on sample size does not vary significantly across these conferences.

WTP for Implementation Information.—But does access to research lead to more effective policies being adopted? At the very end of the beliefs experiment, participants were given the chance to purchase practical information on how to implement ECD policies, using a fresh budget of lottery tickets. We interpret WTP for such advice as a revealed-preference proxy for interest in implementing the policy. Since we experimentally vary bundles of study attributes provided, effect size, developing country context, and large sample, and found that these affect posteriors, we can use these attributes as instruments for participants' posterior beliefs. Online Appendix Table A.8 shows the results. We find that more positive beliefs about ECD programs, shaped through learning about research findings, causally increase WTP for implementation information. This provides clean, experimental evidence on the effect of research information on demand for policy implementation via changed beliefs.

D. *Tests of Motivated Reasoning*

In this section, we test for specific forms of motivated reasoning in belief updating: asymmetric updating and confirmation bias.[26]

Confirmation Bias and Asymmetric Updating.—Confirmation bias is the tendency to acquire and interpret information in a way that confirms one's preexisting beliefs (Nickerson 1998). This phenomenon has been studied in a number of settings, and debates exist as to its prevalence and importance in causing polarization and making individuals immune to evidence (see, e.g., Lord, Ross, and Lepper 1979; Kuhn and Lao 1996; Nyhan and Reifler 2010; Wood and Porter 2019). It is natural to therefore ask if political leaders and other policymakers exhibit confirmation bias when faced with evidence from research on policy effectiveness. Do policymakers who start off with more positive beliefs about a particular policy underreact to negative (disconfirming) information about that policy relative to positive (confirming) information? And do policymakers with negative priors do the reverse? Alternatively, do policymakers systematically respond more to positive information relative to negative information, thereby ending up overoptimistic about policies, as argued by Vivalt and Coville (2020)? We test these hypotheses by estimating equations of the form:

$$(4) \qquad Posterior_{ijs} = \beta_1 Prior_{ij} + \beta_2\big(Signal_{ij} - Prior_{ij}\big) + \beta_3\big(Signal_{ij} - Prior_{ij}\big) \times PositiveSurprise_{ij} + \varepsilon_{ij},$$

where $PositiveSurprise_{ij} = \mathbf{1}\big\{Signal_{ij} - Prior_{ij} > 0\big\}$ is a dummy equal to 1 when the revealed effect-size from the study is larger than the participant's prior, and 0 otherwise. Note, $\beta_3 > 0$ implies that participants place more weight on positive news than on negative news, while $\beta_3 < 0$ would imply placing more weight on negative news than on positive news.[27]

Table 5 reports the results. Column 1 shows that, on average, policymakers do *not* react asymmetrically to positive news relative to negative news. But this comparison may be confounded: the two large-sample studies find smaller effects, and are therefore more likely to lead to negative news. If large-sample studies are given greater weight in updating (as we showed previously), this would tend to counteract any tendency to under-weight bad news. Columns 2 and 3 therefore test for asymmetric updating separately for the large-sample and small-sample studies. The point estimates again do not indicate substantial asymmetric updating, unlike in Vivalt and Coville (2020), although the estimate for large-sample studies in particular is quite imprecise. Column 4 tests for confirmation bias. To do so, we define a variable *ConfirmingNews* equal to 1 if an individual with an above-median prior receives

[26] It is worth noting that we will test for evidence *consistent* with such motivated belief-updating. However, similar empirical patterns could also be generated by Bayesian learners with non-Gaussian priors. Since we largely find null effects, we conclude that we do not observe evidence for motivated reasoning.

[27] See also Vivalt and Coville (2020). An alternative way to set up the estimating equation would be as in equation (3), interacting the prior and signal separately with *PositiveSurprise*. We choose to instead include the *Signal* − *Posterior* term and its single interaction with *PositiveSurprise* for ease of exposition and interpretation.

Table 5—Beliefs Experiment: Testing for Asymmetric Updating and Confirmation Bias

	Posterior			
LHS variable	(1)	(2)	(3)	(4)
Study characteristic	All	Large	Small	All
Prior	0.9957	1.2310	0.9508	1.0165
	(0.0217)	(0.0572)	(0.0236)	(0.0150)
Signal − prior	0.3075	0.6476	0.2429	0.3406
	(0.0499)	(0.1003)	(0.0937)	(0.0235)
Signal − prior × positive surprise	0.0193	0.1166	0.0999	
	(0.0659)	(0.2089)	(0.1077)	
Signal − prior × confirming news				−0.1179
				(0.0573)
Observations	1,131	582	549	1,131
Round	1 and 2	1 and 2	1 and 2	1 and 2
Beliefs about	Municipality	Municipality	Municipality	Municipality
Received study for free	Yes	Yes	Yes	Yes
Clusters (individuals)	731	513	484	731

Notes: OLS results. The dependent variables are posterior beliefs, which are declared after successfully buying the results from a study in each round. *Study characteristic* indicates the sample of studies used in the model (large-sample studies: Colombia and United States; small-sample studies: Michigan and Jamaica; or all studies: Colombia, Jamaica, Michigan, and United States). *Prior* is the belief of the respondent about the effect, right before buying a study. *Signal* is the bought study's effect size. When dealing with a second update in posteriors, the first update is treated as a prior. *Positive surprise* is a dummy which is equal to 1 if the bought study's effect is greater than the respondent's prior about the effect. *Confirming mews* is a dummy which is equal to 1 if the respondent's prior about the effect was above the median (or below the median) and the bought study's effect is greater (smaller) than the respondent's prior, 0 otherwise. In the rows below the coefficients, *Beliefs about* specifies which location the beliefs are elicited for, either the respondent's own municipality or one of the four possible study locations. *Received study for free* indicates whether participant received the information regardless of their WTP. Robust standard errors clustered at the individual level are in parentheses.

a (still more) positive signal or if an individual with a below-median prior receives a (still more) negative signal, and 0 otherwise. The coefficient on the *ConfirmingNews* variable is negative, implying the opposite of confirmation bias. Altogether, we find no evidence for confirmation bias or asymmetric updating on average when policymakers are presented with research evidence on policy effectiveness.

E. *Beliefs Experiment: Discussion*

We have two main findings from the beliefs experiment. First, political leaders in Brazil value learning about research on policy effectiveness. Second, they also change their beliefs when confronted with evidence from research: they place substantial weight on the new information. They place more weight on larger-sample studies, but not on developing-country studies. While attending to sample size indicates a degree of sophistication, we provide suggestive calculations that the sensitivity to sample size is lower than expected from a Bayesian learner.

The experiment has some clear weaknesses. We contrast the effects of a limited number of studies from only three countries. The WTP measure is rather artificial, and comes out of the policymaker's private budget, rather than the likely more-relevant municipal budget, which may have other higher-value uses. We establish effects on beliefs only over a very short period of time and cannot speak to whether the effects

persist. The experiment may also generate demand effects, with participants feeling some social pressure to place weight on the study results (although they completed the experiment privately on a tablet, rather than face-to-face with an experimenter). The policy-adoption experiment described in the next section, in contrast, provides evidence of longer-lasting changes in beliefs and measures effects on actual municipal policy.

III. Policy-Adoption Experiment

In this section, we describe a nationwide field experiment to test whether supplying the heads of local governments with evidence from policy-effectiveness research influences the policies implemented in their polities. We show that informing Brazilian mayors about the effectiveness of a policy to increase tax compliance causally affects not just beliefs, but also adoption of the policy one to two years later.

A. *Background: Taxpayer Reminder Letters*

The essence of our policy-adoption experiment is to inform a treatment group of mayors about the existing research evidence on a particular policy that has been shown to increase tax compliance: reminder letters to taxpayers.

We chose this particular policy for three reasons. First, increasing tax compliance is important to mayors: they reported considerable interest in increasing tax revenues in our focus groups and scoping surveys. Over 90 percent of Brazilian municipalities raise taxes locally and enforcement of municipal taxes is under the control of municipal governments. Like in most developing countries, taxpayer compliance is a challenge in Brazil. A prominent think tank estimates that at least 20 percent of taxpayers do not comply with property taxes, for instance (De Cesare and Smolka 2004).

Second, the effectiveness of reminder letters has been rigorously evaluated in multiple RCTs, including two in Latin America (Coleman 1996; Hasseldine et al. 2007; Del Carpio 2013; Fellner, Sausgruber, and Traxler 2013; Castro and Scartascini 2015; Hallsworth et al. 2017). Such interventions have been found to be surprisingly effective. For instance, Del Carpio (2013) finds that simple reminder letters increased tax compliance in Peru by 10 percent, while letters that additionally included social-norm language by emphasizing that most people pay their taxes on time increased compliance by 20 percent.

Third, reminder letters are inexpensive and relatively easy to implement, while not being obviously politically sensitive. On the one hand, this means that the policy we chose is likely positively selected in terms of the potential for changes in policymaker beliefs to translate into policy change. On the other hand, we expect that reminder letters are likely an effective policy tool in part *because* they are low-cost and easy to implement.

Reminder letters to taxpayers are uncommon but far from unheard of in Brazil. In our endline survey for this experiment, 32 percent of control municipalities reported using some form of reminder messages to taxpayers. This sometimes involved sending letters, but also included other communication channels such as text messages, flyers, and media advertising.

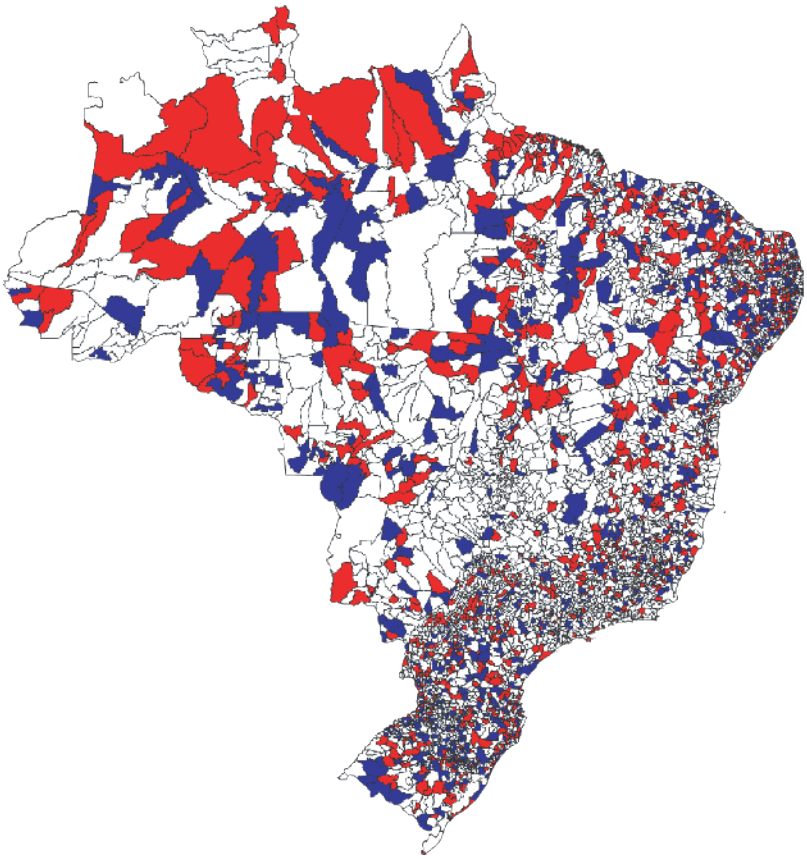

FIGURE 5. POLICY-ADOPTION EXPERIMENT: SAMPLE MUNICIPALITIES

B. *Experimental Setting*

The policy-adoption experiment was conducted at a large CNM convention, the *Novos Gestores* meeting, for recently elected and reelected mayors in October–November 2016. The convention is held every four years to train mayors who are about to start their four-year term the following January. Mayors can attend multiple training sessions led by CNM expert staff on topics ranging from municipal financial planning and budgets to public policy areas, such as urban development, education, health, and tourism. Multiple sessions run in parallel throughout the conference, except for a limited number of plenary sessions. The conference itself ran in stages, with mayors from different regions attending on different days due to capacity constraints. Each mayor attended for two days.

The sample frame for the experiment was mayors attending the convention who represented municipalities with populations between 5,000 and 100,000. The total sample consists of 1,818 municipalities, which represents 45 percent of all municipalities in that population range. Figure 5 shows the spatial distribution of the sample municipalities.

Table 6—Policy-Adoption Experiment: Summary Stats and Balance

	at Baseline			at Endline		
Variables	Mean control	Δ Treatment	*p*-value	Mean control	Δ Treatment	*p*-value
Mayors' characteristics						
Male	88.26	1.41	0.34	90.01	−0.14	0.93
Age	46.76	1.32	0.01	47.08	1.61	0.00
College or more	57.74	−0.76	0.74	57.66	0.73	0.78
Second term	15.69	1.56	0.37	15.18	0.91	0.63
Electoral margin victory	16.73	0.36	0.68	16.61	0.46	0.63
Leftist political party	32.98	2.10	0.35	32.76	1.36	0.58
Municipalities' characteristics						
Population	20.86	−0.06	0.94	20.23	0.06	0.95
College population	5.17	−0.15	0.25	5.47	−0.14	0.31
Public adm. college	32.50	0.89	0.21	33.32	0.25	0.74
Poverty	26.40	−0.27	0.76	23.05	0.11	0.91
Gini	50.33	−0.19	0.54	49.37	0.17	0.61
Big South	51.01	−0.62	0.79	59.92	−2.36	0.36
Per capita income	457.64	3.42	0.75	489.23	2.78	0.81
Local tax revenues (2010–2015)	6.06	0.09	0.68	6.40	0.08	0.75
Joint *F*-test			0.17			0.20
Attrition						
Municipality				19.85	−1.69	0.36
Mayor				48.35	2.28	0.33
Finance staff				24.97	−0.80	0.69

Notes: Sample means of control group and differences in means with respect to treatment group at baseline and endline. There were 937 (751) municipalities in the control group and 881 (721) in the treatment group at baseline (endline). The first block of variables reports characteristics of the mayor that runs the municipality. *Leftist political party* (=1 for mayors belonging to a center-leftist party according to historical political platforms, 0 otherwise). The second block of variables reports characteristics of the municipality. *Population* is the municipality's number of inhabitants (in thousands). *College population* is the municipality's share of adults with college degrees. *Public administration college* is the share of municipal public employees with college degrees. *Poverty* is the municipality's poverty rate. *Gini* is the municipality's Gini coefficient. *Big South* is equal to 1 for municipalities in the south, southeast and midwest regions, 0 otherwise. *Per capita income* is the municipality's monthly income per capita. *Local tax revenues* (*2010–2015*) indicates the average share of municipal tax revenues in total municipal revenues from 2010 to 2015. *Joint significance F*-test, and follow-up survey attrition rate (*Municipality, Mayor, and Finance staff*) at endline.

Table 6 provides descriptive statistics on the sample of mayors and municipalities for the policy-adoption experiment. We see, for example, that almost 90 percent of the mayors are men; about 60 percent have at least a bachelor degree; and 16 percent are in their second and last term in office.[28] The average municipality in the sample has a population of about 21,000 residents.

C. *Experimental Design*

Mayors attending the conference were randomized into treatment ($n = 881$) and control ($n = 937$) groups.[29] All mayors were free to attend any of CNM's regular

[28] This low share of mayors in their second term is explained, in part, by the political crisis Brazil was going through at the time of the most recent municipal elections (2016), which led to a decrease in the proportion of incumbent politicians winning reelection.

[29] The randomization was stratified on the mayor's education level, whether the mayor was term-limited, the average education level among public employees in the municipality, and the municipality's population size, Gini

Novos Gestores training sessions, but only mayors in the treatment group were invited, by email and text message, to attend our research-information sessions. The session was advertised as being on the topic of how to increase local tax revenues, and was framed as a training session organized by CNM as well as researchers at Columbia and Harvard Universities. Since participation was optional, our experiment should be thought of as having an encouragement design. Table 6 shows that the treatment and control groups are largely balanced on mayor's characteristics as well as municipal characteristics.

The information sessions lasted 45 minutes and were led by an experienced local instructor, without foreign researchers present. The instructor began by introducing and defining policy impact, cost-effectiveness, and impact evaluation research. She then provided a description of taxpayer reminder letters, including presenting an example template. Next, she presented the findings (i.e., the estimated effect sizes) of a set of rigorous studies evaluating the impact of taxpayer reminder letters. A list of reminder letter characteristics found to be effective in inducing taxpayers to pay their taxes on time (stating the tax payment deadline; mentioning the possibility of fines and audits for not paying taxes on time; and stating that most people pay their taxes on time) was emphasized, and effect sizes were provided where possible.

The information presented was simplified and the presentation was concise. We avoided jargon and regression tables. The 30-minute presentation was followed by 15 minutes for questions from the audience.[30] At the end of the session, mayors received a professionally produced policy brief with the same information content as the presentation, including references to the cited papers.[31]

The session was offered three to four times during each stage of the conference. Treated mayors could therefore choose to attend when their opportunity cost of time was lowest. Judging from our (unsystematic) field observations, it appeared that the most common counterfactual to attending our information session was networking with other mayors. For other mayors, the opportunity cost was instead attending one of the other simultaneous sessions at the conference. Half of our sessions clashed with a plenary session which taught mayors about municipal finances and budgets and emphasized proper financial planning and fiscal responsibility. The other half clashed with slots during which mayors could have drop-in office hours with the partner organization, or could instead attend a variety of parallel sessions, each of which were themselves offered twice during each stage. We did not clash with other plenary sessions designed around public policy (social policies, urban development policies, and economic policies). No other session at the conference emphasized research information or impact evaluation.

To summarize, attending our information sessions came at the expense of some combination of professional networking, training on municipal budgeting and

coefficient, and region. A slightly larger share of municipalities was assigned to the control group due to logistical concerns associated with our capacity to manage a large number of treatment group participants and the capacity of the room that CNM designated for our intervention.

[30] During the 15 minutes reserved for open discussions with mayors, mayors often asked interesting questions about reminder letters and other alternative policies on tax compliance: for example, whether the effects would be the same if the messages were sent by email or text messages, whether the policy could be used to encourage tax debtors to pay their balance, and whether financial incentives such as discounts or lotteries for paying taxes on time are effective policies. We avoided providing confident answers to such questions.

[31] Online Appendix Section C presents the policy brief.

finances, or sessions on a variety of topics not including impact evaluation, research evidence or economic, social or urban development policies. While our treatment induces greater policy adoption of tax reminder letters, as we will show, it may come at a cost in terms of reduced professional networking or worse knowledge or performance on a diffuse range of outcomes we do not observe. This does not change our conclusion, however, that providing research information did lead to a change in policy.

D. *Data*

To measure how the research information provision affected political leaders' beliefs and ultimate policy adoption, we conducted in-depth phone surveys of relevant municipal officials from treatment and control municipalities 15 to 24 months after the session. We attempted to reach the bureaucrat in charge of implementing tax policy in each municipality, as well as the mayors themselves.[32]

The survey was supervised by a research assistant, and conducted by a team of nine surveyors who were blinded to treatment status and the research hypotheses. When the survey ended after 10 months of phone calls, we had successfully interviewed at least one person in 81 percent of our sample of municipalities: 75 percent of the chief tax bureaucrats and 51 percent of the mayors in the sample. We were not able to make any contact with 10 percent of the sample municipalities, due to not being able to locate a working phone number. This share was also balanced across treatment and control groups.[33] There was no differential attrition between treatment and control groups, and observable characteristics of the successfully contacted municipalities are similar across both groups, as reported in Table 6.

The survey lasted approximately 15 minutes. The key outcomes asked about whether the municipality sends reminders to taxpayers to pay their taxes, and whether the messages feature the characteristics described in the information session and evaluated in the literature: the due date, the possibility of fines or audits, and language regarding the social norm of paying taxes on time. An important secondary outcome measured in the survey was beliefs about policy effectiveness. Specifically, we elicited quantitative beliefs about the likely impact of such a policy, in terms of percent changes in tax compliance, even if the municipality reported not using such reminders. In addition, we asked questions that served as attention and comprehension checks as well as questions about one potential policy substitute to reminder letters (namely financial incentives for tax payers) and placebo questions on which we would expect null effects of the treatment (the use of e-procurement platforms).

In addition to the phone survey, we gathered demographic, electoral, and budgetary data from official sources for all municipalities for which such data are

[32] Typically, secretaries of finance are responsible for the tax division in Brazilian municipalities. Nevertheless, we specifically asked municipalities' telephone attendants to pass the call on to the person in charge of the tax division. Once we were transferred, we confirmed whether the person actually held that position or asked to get the phone number of the person in charge of implementing tax policy.

[33] On average, many hours of work were needed before we could talk to the chief tax bureaucrats and mayors over the phone, mainly collecting municipalities' phone numbers. Not all Brazilian municipalities publish or have updated contact information on their websites, so we collected phone numbers through Google searches, Facebook, by calling other local institutions such as hospitals and schools, etc.

available.[34] It is not possible to observe tax compliance itself in the administrative data so our primary outcome is whether municipalities implemented the policy. Since concerns about experimenter demand effects or other reporting biases may arise for reports from mayors, we separately report responses from tax department bureaucrats and mayors.

E. *Results*

Participation in Information Sessions.—Of the mayors in the treatment group, 37.9 percent chose to attend our session. In contrast, less than 1 percent of control group mayors attended the session. The opportunity costs of attending, forgoing the opportunity to attend other parallel training sessions or conducting meetings with other politicians and officials, were meaningful, although difficult to quantify. Moreover, some mayors did not have accurate contact information stored in the CNM system, and thus did not receive our invitation messages at all. We therefore consider 37.9 percent to be a fairly high rate of treatment group participation.

Online Appendix Table A.9 reports predictors of participation in the research-information session. Younger and college-educated mayors are 7 and 15 percentage points more likely to participate than others, but term-limited mayors are no less likely to participate than mayors in their first term. None of the municipal characteristics, such as poverty rates, inequality, or income per capita, predict participation.

Policy Adoption.—We find that a mayor attending the research-information session leads to about a 10 to 11 percentage point increase in the use of reminder letters to taxpayers, an increase of over 33 percent over the proportion of control group municipalities that had started using such a reminder policy at some point in the past. Table 7 presents Treatment-on-Treated (ToT) estimates, using randomized treatment status as an instrument for participation in the information session.[35] The outcome variable is a dummy equal to 1 if the respondent reports that the carefully described policy is used in their municipality and 0 otherwise. Standard errors are clustered at the municipality level. In Column 1, the ToT coefficient is 10.3 percentage points (SE 5.3 percentage points), compared to a base of 31.7 percentage points in the control group. Column 2 drops respondents who failed an attention check, leaving the coefficient largely unchanged.[36] Most importantly, the point estimates are very similar if we restrict attention to responses from mayors (column 3) or tax department officials (column 4). Given that we have little concern about tax department officials misreporting details of tax compliance policies differentially between treatment and

[34] See data sources for demographic and electoral data in footnote 15. Budgetary data were retrieved from the National Treasury, which compiles and releases self-reported accounting records from all Brazilian municipalities every year (FINBRA 2010–2012, 2013–2015).

[35] One possible violation of the exclusion restriction in the IV estimates is that treatment-group mayors who did not attend an information session were later emailed a link to the policy brief. Relatively few of these links were clicked on, but it is likely that at least some additional treatment-group mayors read the brief without attending a session. Online Appendix Table A.11 therefore presents intent-to-treat estimates.

[36] The attention check was: "The tax reminders sent informed taxpayers that the Brazilian constitution was reformed in 1988." Since we consider this exceedingly unlikely as text for a tax reminder, we infer that respondents who answer "yes" to this question are simply not paying attention or following the questions.

Table 7—Policy-Adoption Experiment: Policy Adoption: Tax Reminders

	Adopted			
LHS variable	(1)	(2)	(3)	(4)
Information session	0.1031	0.0934	0.1125	0.1028
	(0.0531)	(0.0551)	(0.0791)	(0.0656)
Observations	2,271	2,055	913	1,358
Respondent	All	All	Mayor	Finance staff
Drops inattentive	No	Yes	No	No
Clusters (municipalities)	1,465	1,413	913	1,358
Mean control	0.317	0.298	0.367	0.283

Notes: 2SLS results. The dependent variable is a dummy which is equal to 1 if respondent says the policy was adopted in municipality, 0 otherwise. *Information session* is a dummy which is equal to 1 if the municipality's mayor attended the information session about tax reminders, 0 otherwise. This last variable is instrumented with treatment assignment. In the rows below the coefficients, *Drops inattentive* refers to whether respondents who failed the survey attention check component of the reminders policy are excluded from the model, where the attention check was "The tax reminders sent informed taxpayers that the Brazilian constitution was reformed in 1988." Robust standard errors clustered at the municipality level are in parentheses.

control groups, this increases our confidence that the effects we estimate are not driven by reporting biases.[37]

Online Appendix Tables A.12 and A.13 report an exploratory analysis of heterogeneity in treatment effects on policy adoption by mayoral and municipal characteristics. No clear evidence of heterogeneity emerges, partly due to limited statistical power. Only one of fourteen tests (heterogeneity by income inequality) is significant at the 10 percent level, with more unequal municipalities responding less to the treatment. Term-limited mayors are not substantially less likely to adopt reminder letters. The point estimates suggest that mayors of above median age and suggest that mayors of above-median age and victory margins are less likely to adopt reminder letters, while leftist mayors are more likely to adopt, but none of these estimates are statistically significant even without correcting for multiple hypotheses testing. Online Appendix Table A.14 reports effects separately for the different design components of taxpayer reminder letters, and shows that the effects are fairly similar on the probability of using letters emphasizing the due date, mentioning the threat of audits/penalties, and mentioning social norm language, although the latter is a larger effect in relative terms, since it is particularly unlikely to be used in the control municipalities. Finally, online Appendix Table A.15 reports no effects on a placebo question (the use of e-procurement in municipal government), and reports no effects on the use of financial incentives for compliance with taxes, a common policy which might conceivably have been seen as a substitute for the reminder-letters policy.

Beliefs.—We also measured beliefs about the effectiveness of reminder letters, which, especially given the evidence presented in Section II, is a plausible mechanism through which the ultimate impact on policy adoption may arise. We asked respondents about the likely effect of the policy in their municipality, whether or

[37] Online Appendix Table A.10 reports estimates which include controls. Once again, the point estimates are almost identical.

Table 8—Policy-Adoption Experiment: Accuracy of Beliefs and Policy Adoption: Tax Reminders

	(1)	(2)	(3)	(4)
Panel A. LHS variable: Accuracy of beliefs				
Information session	1.3975	1.5416	1.2463	1.5192
	(0.5209)	(0.5573)	(0.7324)	(0.6868)
Mean control	−6.980	−6.999	−6.859	−7.062
Panel B. LHS variable: Adopted				
Accuracy of beliefs	0.0856	0.0718	0.1207	0.0709
	(0.0500)	(0.0449)	(0.0960)	(0.0538)
Mean control	0.310	0.289	0.361	0.275
Observations	2,172	1,963	857	1,315
Respondent	All	All	Mayor	Finance staff
Drops inattentive	No	Yes	No	No
Clusters (municipalities)	1,434	1,378	857	1,315

Notes: 2SLS results where treatment assignment is the instrument for *Information session* (in panel A) and for *Accuracy of beliefs* (in panel B). In panel A, the dependent variable—Accuracy of beliefs—is the absolute difference multiply by −1 between self-reported beliefs about effect sizes of tax reminders on local tax revenues and the 12 percent informed effect size of the reminder letters policy during the information session. *Information session* is a dummy which is equal to 1 if the municipality's mayor attended the information session about tax reminders, 0 otherwise. In panel B, the dependent variable is a dummy which is equal to 1 if respondent says the policy was adopted in municipality, 0 otherwise. *Accuracy of beliefs* is the absolute difference multiplied by −1 between self-reported beliefs about effect sizes of tax reminders on local tax revenues and the 12 percent informed effect size of the reminder letters policy during the information session. In the rows below the coefficients of the last panel, *Drops inattentive* refers to whether respondents that failed the survey attention check component of the reminders policy are excluded from the model, where the attention check was "The tax reminders sent informed taxpayers that the Brazilian constitution was reformed in 1988." Robust standard errors clustered at the municipality level are in parentheses.

not the policy was currently implemented. We compare their stated beliefs with the main estimated effect size of 12 percent shared with participating mayors in the research-information session and policy briefs provided. Unlike in our beliefs experiment, it was not possible to provide benchmarks and comprehension checks during the short phone survey, so the measures must necessarily be treated with some caution. Panel A of Table 8 shows that attending the information session, instrumented using treatment assignment, increased the "accuracy" of beliefs even 15–24 months after treatment. Specifically, the absolute deviation of beliefs from the effect size mentioned in the research-information session is 20 percent lower than in the control group. Comparing columns 3 and 4 reveals that beliefs became more accurate not just among mayors, but also among tax-department bureaucrats, implying information-flow within the municipal government. This was perhaps made easier by providing the participating mayors with shareable policy briefs.

Panel B of Table 8 instead estimates the effect of belief accuracy on policy adoption, now instrumenting for belief accuracy using treatment assignment. Of course, this requires making the debatable assumption that the treatment only affects adoption through beliefs. The estimates imply that increasing belief accuracy by 1 percentage point (i.e., reducing the absolute deviation by 1 pp on a base of about 7 pp) increases adoption by 8 percentage points (SE 5 pp). Of course, the effects may operate also

through other channels such confidence, salience, etc., as discussed below. These magnitudes must therefore be treated as descriptive. It is also worth noting that the relevant beliefs were presumably those at the time of the policy-adoption decision, which we do not observe.[38]

F. *Policy-Adoption Experiment: Discussion*

This experiment has one simple but important result: when political leaders in Brazil are provided information from research on the impact of a cost-effective policy, they change the actual policies in use in their polities. This implies that, consistent with the findings from our beliefs experiment, policymakers are open to new evidence; care about policy effectiveness; and have at least some capacity and desire to translate evidence into policy change.

Some caveats to this interpretation are worth noting. First, we cannot rule out that the estimated effects are driven in part by mayors simply learning of the existence of taxpayer reminder policies, rather than due to the quantitative estimates of their impact from research. As noted above, however, taxpayer reminder policies are far from unknown in Brazil, with about one-third of municipalities already using some form of such reminders. We also found evidence of more accurate beliefs in the treatment group, although we cannot rule out that effects would have been similar had we simply provided a policy recommendation stripped of any evidence. Second, we considered a policy that is inexpensive and relatively easy to implement. Other effective policies may have higher up-front costs, be more technically demanding, or be more politically sensitive, in which case changing beliefs about effectiveness may not as readily translate into policy change. Third, we estimate the effect of providing research information in a particular context: an information session designed by researchers at reputed foreign universities, at a conference organized by a trusted local organization. Research findings received from other sources, such as local think tanks, academics, or media sources, may be differently received. Similarly, policymakers seeking to unearth relevant research information themselves may have difficulty finding and interpreting relevant and high-quality information. On the other hand, our policy adoption experiment also does not capture the numerous, less direct channels through which research may ultimately influence policy practice.

IV. Conclusion

Policy is important for economic development. What role can policy-effectiveness research play in spurring the spread of effective policies and the abandonment of ineffective ones? One possibility is that lack of (access to) research information is not a binding constraint on policy choice, for example because political leaders are self-interested and electoral competitive pressures too weak to motivate the effort required to change policy, or because leaders have limited real power over the policies in use. Alternatively, frictions may constrain political leaders' access to existing research.

[38] Online Appendix Table A.16 reports estimates which include controls, with very similar results.

In this paper, we investigate how informing political leaders about research findings affects policy beliefs and practice. Using experiments with the elected heads of Brazil's local governments, mayors, we first show that political leaders value access to impact evaluations, and update their beliefs when informed of the research findings. Mayors (and other local policymakers in our sample) appear to be fairly sophisticated consumers of accessible research, for example paying more for studies, such as those with a large sample size, that subsequently affect their beliefs more. In the second half of the paper, we show that providing mayors with research findings documenting positive impact of an inexpensive and easy-to-implement policy increases the probability that their municipality implements the policy by 10 percentage points. Making research information directly and easily available to policymakers therefore appears to influence policy. This suggests that information frictions may play an important role in explaining failures to adopt policies which have been proven to be effective.

It is arguably surprising that such information frictions persist. After all, even if political leaders themselves do not read academic journals, information frictions should generate incentives for actors interested in enhancing social welfare to access academic research and connect policy research with practice to eliminate these frictions. Empirically, the reach of think tanks and other organizations that institutionalize and scale up transmission of research findings to political leaders still appears limited in developing countries. Moreover, policymakers might face the problem of information overload, with numerous motivated actors attempting to persuade them by providing them with selective pieces of evidence and information. We hope that future research will expand our understanding of how research's impact on policy practice can be better understood and enhanced.

REFERENCES

Akhtari, Mitra, Diana Moreira, and Laura Trucco. 2018. "Political Turnover, Bureaucratic Turnover, and the Quality of Public Services." Unpublished.

AtlasBR. 2010. "Censo demográfico." *Atlas do Desenvolvimento Humano no Brasil*. http://www.atlasbrasil.org.br/acervo/biblioteca (accessed October 2016).

Attanasio, Orazio P., Camila Fernández, Emla O. A. Fitzsimons, Sally M. Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. 2014. "Using the Infrastructure of a Conditional Cash Transfer Program to Deliver a Scalable Integrated Early Child Development Program in Colombia: Cluster Randomized Controlled Trial." *BMJ* 349: g5785.

Banuri, Sheheryar, Stefan Dercon, and Varun Gauri. 2019. "Biased Policy Professionals." *World Bank Economic Review* 33 (2): 310–27.

Barnett, W. S. 2011. "Effectiveness of Early Educational Intervention." *Science* 333 (6045): 975–78.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India." *Science* 335 (6068): 582–86.

Benjamin, Daniel J. 2019. "Errors in Probabilistic Reasoning and Judgment Biases." In *Handbook of Behavioral Economics: Applications and Foundations*, Vol. 2, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 69–186. Amsterdam: Elsevier.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu. 2020. "The Glittering Prizes: Career Incentives and Bureaucrat Performance." *Review of Economic Studies* 87 (2): 626–55.

Besley, Timothy, Jose G. Montalvo, and Marta Reynal-Querol. 2011. "Do Educated Leaders Matter?" *Economic Journal* 121 (554): F205–27.

Best, Michael Carlos, Jonas Hjort, and David Szakonyi. 2018. "Individuals and Organizations as Sources of State Effectiveness, and Consequences for Policy Design." NBER Working Paper 23350.

Beynon, Penelope, Christelle Chapoy, Marie Gaarder, and Edoardo Masset. 2012. "What Difference Does a Policy Brief Make?" Unpublished.

Beynon, Penelope, Christelle Chapoy, Marie Gaarder, and Edoardo Masset. 2013. "What Is the Impact of a Policy Brief? Results of an Experiment in Research Dissemination." *Journal of Development Effectiveness* 5 (1): 50–63.

Bidwell, Kelly, Katherine Casey, and Rachel Glennerster. 2020. "Debates: Voting and Expenditure Responses to Political Communication." *Journal of Political Economy* 128 (8): 2880–2924.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics* 128 (1): 1–51.

Carabetta, João, Ricardo Dahis, Fred Israel, and Fernanda Scovino. 2020. Data Basis: Open Data Repository at https://www.basedosdados.org/dataset/br-bd-diretorios-brasil/resource/c1deb363-ffba-4b1e-95dc-c5e08311852e (accessed October 2018).

Castro, Lucio, and Carlos Scartascini. 2015. "Tax Compliance and Enforcement in the Pampas Evidence from a Field Experiment." *Journal of Economic Behavior and Organization* 116: 65–82.

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409–43.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. 2016. "Decision Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires." *Quarterly Journal of Economics* 131 (3): 1181–242.

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.

Coleman, Stephen. 1996. "The Minnesota Income Tax Compliance Experiment: State Tax Results." Unpublished.

De Cesare, Claudio, and Martim Smolka. 2004. "Diagnóstico sobre o IPTU." Lincoln Institute of Land Policy.

Del Carpio, Lucia. 2013. "Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru." Unpublished.

DellaVigna, Stefano, and Devin Pope. 2018. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy* 126 (6): 2410–56.

Dizon-Ross, Rebecca. 2019. "Parents' Beliefs about Their Children's Academic Ability: Implications for Educational Investments." *American Economic Review* 109 (8): 2728–65.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128 (4): 1499–545.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2018. "The Value of Regulatory Discretion: Estimates from Environmental Inspections in India." *Econometrica* 86 (6): 2123–60.

Easterly, William, and Steven Pennings. 2017. "Shrinking Dictators: How Much Economic Growth Can We Attribute to National Leaders?" DRI Working Paper 94.

Eil, David, and Justin M. Rao. 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics* 3 (2): 114–38.

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): 634–60.

Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123 (2): 703–45.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. 2017. "The Personnel Economics of the Developing State." In *Handbook of Economic Field Experiments*. Vol. 2, edited by Abhijit Vinayak Banerjee and Esther Duflo, 467–514. Amsterdam: Elsevier.

FINBRA. 2010–2012. "Dados contábeis dos Municípios." *Tesouro Nacional.* https://www.tesourotransparente.gov.br/publicacoes/finbra-dados-contabeis-dos-municipios-1989-a-2012/2012/26 (accessed October 2016).

FINBRA. 2013–2015. "Dados contábeis dos Municípios." *Tesouro Nacional.* https://siconfi.tesouro.gov.br/siconfi/pages/public/consulta_finbra/finbra_list.jsf (accessed October 2016).

Firpo, Sergio, Renan Pieri, and André Portela Souza. 2017. "Electoral Impacts of Uncovering Public School Quality: Evidence from Brazilian Municipalities." *EconomiA* 18 (1): 1–17.

Fujiwara, Thomas, and Leonard Wantchekon. 2013. "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin." *American Economic Journal: Applied Economics* 5 (4): 241–55.

Grantham-McGregor, Sally M., Christine A. Powell, Susan P. Walker, and John H. Himes. 1991. "Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study." *Lancet* 338 (8758): 1–5.

Griffin, Dale, and Amos Tversky. 1992. "The Weighing of Evidence and the Determinants of Confidence." *Cognitive Psychology* 24 (3): 411–35.

Grigorieff, Alexis, Christopher Roth, and Diego Ubfal. 2020. "Does Information Change Attitudes towards Immigrants? Representative Evidence from Survey Experiments." *Demography* 57: 1117–43.

Haaland, Ingar, and Christopher Roth. 2020. "Labor Market Concerns and Support for Immigration." *Journal of Public Economics* 191: DOI: 104256.

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.

Hasseldine, John, Peggy Hite, Simon James, and Marika Toumi. 2007. "Persuasive Communications: Tax Compliance Enforcement Strategies for Sole Proprietors." *Contemporary Accounting Research/Recherche Comptable Contemporaine* 24 (1): 171–94.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini. 2021. "Replication Data for: How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E122661V1.

Hoffmann, Vivian, Pamela Jakiela, Michael Kremer, Ryan Sheely, and Matthew Goodkin-Gold. 2017. "There Is No Place Like Home: Theory and Evidence on Decentralization and Politician Preferences." Unpublished.

IBGE. 2016. "Estimativas da população." Instituto Brasileiro de Geografia e Estatística. https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=downloads (accessed October 2016).

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125 (2): 515–48.

Jones, Benjamin F., and Benjamin A. Olken. 2005. "Do Leaders Matter? National Leadership and Growth since World War II." *Quarterly Journal of Economics* 120 (3): 835–64.

Kahneman, Daniel, and Amos Tversky. 1972. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* 3 (3): 430–54.

Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." *Quarterly Journal of Economics* 131 (1): 219–71.

Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken. 2019. "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings." *American Economic Review* 109 (1): 237–70.

Kling, Jeffrey R., Sendhil Mullainathan, Eldar Shafir, Lee C. Vermeulen, and Marian V. Wrobel. 2012. "Comparison Friction: Experimental Evidence from Medicare Drug Plans." *Quarterly Journal of Economics* 127 (1): 199–235.

Kuhn, Deanna, and Joseph Lao. 1996. "Effects of Evidence on Attitudes: Is Polarization the Norm?" *Psychological Science* 7 (2): 115–20.

Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37 (11): 2098–109.

Malmendier, Ulrike, Stefan Nagel, and Zhen Yan. 2017. "The Making of Hawks and Doves: Inflation Experiences on the FOMC." NBER Working Paper 23228.

Martinez-Bravo, Monica. 2014. "The Role of Local Officials in New Democracies: Evidence from Indonesia." *American Economic Review* 104 (4): 1244–87.

Martinez-Bravo, Monica. 2017. "The Local Political Economy Effects of School Construction in Indonesia." *American Economic Journal: Applied Economics* 9 (2): 256–89.

Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 2011. "Managing Self-Confidence: Theory and Experimental Evidence." NBER Working Paper 17014.

Mulas-Granados, Carlos. 2017. "How to Fix the Fiscal Crisis in Brazil's States." International Monetary Fund.

Nath, Anusha. 2015. "Bureaucrats and Politicians: How Does Electoral Competition Affect Bureaucratic Performance?" IED Working Paper 269.

Nellis, Gareth, Thad Dunning, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, and Catlan Reardon. 2019. "Learning about Cumulative Learning: An Experiment with Policy Practitioners." Unpublished.

Nickerson, Raymond S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2 (2): 175–220.

Nyhan, Brendan, and Jason Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32 (2): 303–30.

Persson, Torsten, and Guido Tabellini. 2002. *Political Economics: Explaining Economic Policy.* Vol. 1. 1st ed. Cambridge, MA: MIT Press.

Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, Gary Shapiro, Pam Broene, Frank Jenkins, et al. 2010. *Head Start Impact Study: Final Report*. US Department of Health and Human Services, Administration for Children & Families.

RAIS. 2012–2016. "Microdados RAIS." *Ministério do Trabalho.* ftp://ftp.mtps.gov.br/pdet/microdados/RAIS/ (accessed October 2016).

Rasul, Imran, and Daniel Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *Economic Journal* 128 (608): 413–46.

Rogger, Daniel, and Ravi Somani. 2019. *Hierarchy and Information*. Washington, DC: The World Bank.

Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Nores. 2005. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40.* Ypsilanti, MI: High/Scope Press.

Terra, Rafael, and Enlinson Mattos. 2017. "Accountability and Yardstick Competition in the Public Provision of Education." *Journal of Urban Economics* 99: 15–30.

TSE. 2012 and 2016. "Repositório de dados eleitorais." *Tribunal Superior Eleitoral.* https://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1/repositorio-de-dados-eleitorais (accessed October 2016 and May 2018).

Vivalt, Eva, and Aidan Coville. 2020. "How Do Policymakers Update?" Unpublished.

Walker, Susan P., Susan M. Chang, Christine A. Powell, and Sally M. Grantham-McGregor. 2005. "Effects of Early Childhood Psychosocial Stimulation and Nutritional Supplementation on Cognition and Education in Growth-Stunted Jamaican Children: Prospective Cohort Study." *Lancet* 366 (9499): 1804–807.

Wood, Thomas, and Ethan Porter. 2019. "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence." *Political Behavior* 41 (1): 135–63.

Xu, Guo. 2018. "The Costs of Patronage: Evidence from the British Empire." *American Economic Review* 108 (11): 3170–98.

Yao, Yang, and Muyang Zhang. 2015. "Subnational Leaders and Economic Growth: Evidence from Chinese Cities." *Journal of Economic Growth* 20 (4): 405–36.