Accountability in Time:
Evolution and Expertise in Participatory Institutions

Brian Palmer-Rubin[1]
Jésica E. Tapia Reyes[2]
Daniel Berliner[3]
Aaron Erlich[4]
Benjamin E. Bagozzi[5]

November 1, 2024

Forthcoming, *World Politics*

Abstract:

How do participatory institutions change over time? While previous research has focused on exogenous changes such as legal reform or leadership replacement, institutions also evolve endogenously through processes of behavioral and compositional change, on the parts of citizen claimants and government officials. These processes can gradually reshape institutions to become more responsive to either expert or non-expert claimants—which we refer to as *brokered* and *grassroots* models of social accountability. In the context of Mexico's access-to-information system, we analyze nearly two million information requests and responses filed between 2003 and 2019, using new machine-learning based measures. We find evidence of claimants becoming more sophisticated over time and officials becoming more responsive to these expert claimants, both consistent with a brokered accountability model. Both quantitative and qualitative evidence reveal mechanisms of behavioral and compositional change by citizen claimants and government agents.

Keywords: participation, accountability, access-to-information, institutional change, Mexico,

I. Introduction

How do participatory institutions change over time? The past few decades have witnessed a wave of research on participatory institutions of social accountability designed to involve citizens in governance and oversight. These studies highlight exogenous shifts such as legal reforms and leadership change, which led to the success—and later demise—of such institutions

[1] Associate Professor of Political Science and International Relations, University of Southern California, Los Angeles, California; Corresponding author. Email: brian.pr@usc.edu
[2] Independent Researcher, Mexico City
[3] Professor of Political Science and Public Policy, Department of Government, London School of Economics, London
[4] Associate Professor of Political Science, McGill University, Montreal, Québec
[5] Associate Professor of Political Science and International Relations, University of Delaware, Newark, Delaware

as Guatemala's anti-corruption commission[6], policy councils in Brazil[7], and Madrid's municipal participation platform.[8] However, more gradual *endogenous* processes can also lead to shifts in function for these institutions—such as increasing orientation toward claims brought by different classes of citizen claimants. Our study was inspired by conversations with Mexican activists, who complained that Mexico's celebrated access-to-information (ATI) system has become increasingly difficult to navigate for lay citizens during its two-decade history.[9]

ATI systems belong to a class of institutions designed for social accountability, "an approach towards building accountability that relies on civic engagement."[10] Many accountability institutions around the world are similarly participatory: including citizen policy councils,[11] participatory budgeting,[12] local assemblies,[13] independent monitoring organizations,[14] citizen anti-corruption bodies,[15] complaint mechanisms,[16] public reporting platforms,[17] and ombudspersons.[18] Defining traits of social-accountability institutions are that they: (1) feature repeated interaction between ordinary citizens and state actors; and (2) are oriented towards improving government accountability. Excluded from this category are participatory institutions with no accountability orientation at all, such as pure co-production, and elite-driven or "horizontal" (O'Donnell 1998) accountability institutions, such as auditing agencies, which lack regular engagement with citizens.[19]

---

[6] Call and Hallock 2020
[7] Donaghy 2020
[8] Olaizola and Grasso 2023
[9] Parallel to this article, we created a policy brief oriented to Mexican civil society and government actors (Berliner et al. 2022). Also see the related interactive website: https://bigdataytransparenciamx.lse.ac.uk/
[10] Malena, Forster, and Singh 2004, 3
[11] Donaghy 2020; Mayka 2019
[12] Wampler 2015
[13] Sanyal and Rao 2019
[14] Kosack, Tolmie, and Griffin 2010; Robinson 2006
[15] Baniamin and Jamil 2018; Shim and Eom 2008
[16] Kruks-Wisner 2021
[17] Buntaine, Hunnicutt, and Komakech 2021
[18] Wille and Bovens 2022
[19] The concept of "accountability" is highly contested and often subject to conceptual stretching. Certain approaches delineate between institutions of transparency and accountability (Schedler 1999; Fox 2007), with ATI systems following under the former category—in terms of enabling "answerability," but not directly enabling citizen sanctioning of politicians. Others categorize ATI systems as institutions of "diagonal" accountability, given that citizen claimants typically rely on state oversight bodies for sanctioning (Lührmann, Marquardt, and Mechkova 2020). Our situating of request-based ATI systems as a type of social-accountability institution relies on the core role of citizen request-making, the legal mandate of government agents to respond, and the potential for disclosed information to enable formal or informal sanction by government *or* societal actors.

Such social-accountability institutions are potentially powerful tools for improving governance while fomenting civic engagement.[20] Widely promoted by the international development community and bolstered by advances in digital technology that lower barriers to access, such institutions have spread rapidly. To wit, the number of countries with ATI systems has expanded from roughly 20 in 1990 to over 100 today.[21] Yet, social-accountability institutions can be difficult to sustain given their dependence on complex forms of engagement by a broad variety of actors.[22] Furthermore—and most relevant for our purposes—such institutions are subject to "mission drift"[23] because they involve high levels of discretion in implementation and frequent interactions between politicians and citizens, during which each has opportunities to learn and adapt their behavior.

Of particular relevance are cases in which such institutions evolve to become more responsive to elites. While ATI systems and other social accountability institutions purport to engender relationships of accountability between citizens and the state, ordinary citizens often face significant obstacles to make these systems work for their purposes. In contrast, economic elites enjoy the means to contract lawyers and other experts to navigate institutional channels. While existing literature has addressed the aspects of institutional *design* that lend themselves to broad citizen empowerment,[24] we focus on informal and gradual transformations that are more difficult to observe. Our goal is, thus, to uncover the endogenous drivers of these kinds of institutional evolution.

Our dependent variable–*responsiveness*—takes on a double connotation for ATI systems. At the micro-level, responsiveness refers to the degree to which citizen claims receive actual responses from government agents. Thus, responsiveness to an information request is reflected in the timeliness and quality of information provided.[25] At the macro-level, responsiveness reflects whether the institution lends itself to accountability-seeking forms of political activity, therein enhancing democratic representation.[26] At either of these levels, responsiveness can vary across categories of claimants, such as elite civil society actors, grassroots activists, private firms, or lay citizens.

Thus, a first contribution is to build a framework to understand differential responsiveness, distinguishing between expert and non-expert claims. Existing research is often preoccupied with the binary question of *whether* a given institution of accountability yields responsiveness. Such studies typically focus *either* on processes involving individual lay citizens[27] or—more rarely—

---

[20] Ahmad et al. 2003; Smulovitz and Peruzzotti 2000
[21] UNESCO 2022
[22] Fox 2015; McGee and Gaventa 2011
[23] Ebrahim, Battilana, and Mair 2014
[24] Arkedis et al. 2021; Fox 2015; Hossain, Joshi, and Pande 2024; Lieberman, Posner, and Tsai 2014
[25] Berliner et al. 2021; Fung 2013
[26] Schedler 1999
[27] Banerjee et al. 2010; Besley and Burgess 2002; Dunning and others 2019; Ferraz and Finan 2008; Lieberman, Posner, and Tsai 2014

scaled up demand-making by civil society organizations.[28] However, institutions may prove to be more responsive to some types of claimants than to others. Our approach thus foregrounds the question of institutional bias in responsiveness. We distinguish between a *grassroots* model of accountability, where institutions are most useful for general citizens' routine needs and a *brokered* model, where the institution is most useful for experts. By analyzing ATI, a type of institution designed to serve both lay and expert claimants, we elucidate the process by which Mexico's ATI system transformed to favor the latter over the former.

A second contribution—both theoretical and methodological—is to diagnose the *gradual institutional change* that unfolds over time in institutions of accountability. Until now, research on information and accountability has typically focused on short time frames, such as the above-cited experimental evaluations of informational treatments involving individual citizens. Some of these studies have noted a feedback loop in which citizen claim making activity not only enhances—but is also a product of—government responsiveness.[29] However, under-acknowledged are the gradual endogenous processes that produce transformations in how these institutions function over extended periods of time. By incorporating insights from historical institutionalism,[30] we develop a new framework for understanding gradual change in social-accountability institutions. In so doing, we build on an incipient literature that analyzes gradual processes of institutionalization and conversion in institutions of accountability in developing and non-democracies.[31] We also build on recent work demonstrating endogenous change in non-state online institutions,[32] both with a broader set of theoretical mechanisms and an application to formal governmental institutions.

We identify and empirically demonstrate three mechanisms of endogenous change, corresponding to government *obligants* (those charged with responding to citizen claims) and citizen *claimants* (those making claims on the state). First, our findings reinforce previous research detailing changes on the *supply side* of accountability: the behavior of obligants who develop strategies to evade responsiveness or reduce workloads.[33] More novel, we additionally detail two modes of *demand-side* change: the changing composition of citizen claimants—attrition or intensification by distinct groups; and behavioral change (or "learning") by some claimants who adapt strategies to navigate the institution. In terms of Mahoney and Thelen,[34] these three mechanisms interact to create a self-reinforcing process of "conversion," in which the

[28] Crepaz 2020; Fox 2015; Gaventa and Barrett 2012; Goetz and Jenkins 2001
[29] Buntaine, Nielson, and Skaggs 2021; Dipoppa and Grossman 2020; Grossman, Platas, and Rodden 2018; Sjoberg, Mellon, and Peixoto 2017
[30] Mahoney and Thelen 2010; Pierson 2011
[31] Falleti and Riofrancos 2018; Kim et al. 2021; Michener, Contreras, and Niskier 2018; Relly et al. 2020
[32] Steinsson 2024
[33] Berliner et al. 2021; Distelhorst and Hou 2017; Kim et al. 2021; Mungiu-Pippidi 2015
[34] Mahoney and Thelen 2010

informal rules of the institution evolve given high levels of discretion in interpretation and enforcement.[35]

Our approach combines quantitative analysis of textual citizen claims with qualitative evidence of mechanisms driving institutional change. Our main analytical effort is based on all federal-level information requests submitted to Mexico's ATI system from its inception in 2003 until 2019, and their corresponding responses. We use machine learning based on a hand-coded sample of nearly 5,000 requests to measure the full universe of traits such as the expertise of the requester, the theme of information, and the quality of the response. We analyze these to elucidate over-time change as citizen claimants and government obligants react to each other. We pair these statistical approaches with qualitative evidence enabling us to better understand changes among both civil society claimants and agency personnel. Our approach is inspired by the increasing recognition that strong institutions of accountability are not simply a matter of "getting the incentives right," but are rather embedded in complex equilibria conditioned by the behavior of diverse actors.[36]

Our focus is Mexico, which in 2003 established a federal ATI institution and online request-response portal. The 2002 federal law underpinning these innovations has been widely hailed as one of the strongest in the world.[37] Prior research demonstrates that Mexican citizens use this pioneering system for both private needs and the pursuit of broader public goods such as rooting out government corruption.[38] Yet it is also clear that government officials have incentives to be *non-responsive* to certain classes of citizen requests due to electoral motives, bureaucratic constraints, and agency discretion.[39] As the transitional democracy with the longest-lived and perhaps most institutionalized ATI system, we consider Mexico to be a crucial case. We would expect these processes of gradual change to be even more pronounced in systems that are less institutionalized,[40] such as ATI systems in South American countries.[41]

Our quantitative findings confirm that Mexico's ATI system has evolved to be significantly more responsive to expert claimants. For instance, we found that in the earliest years of the system, otherwise-typical "non-expert" requests (one standard deviation below the mean on our expertise score) received responses three days faster than otherwise-typical "expert" requests (one standard deviation above the mean). By the end of our sample (2019), non-expert requests were receiving responses two days *slower* on average. Further analyses explore the mechanisms behind this evolution, signaling both demand-side and supply-side processes that drive an increasing expert bias. On the demand side, we find that an increasing portion of requests exhibit

---

[35] While the present analysis does not address it empirically, social-accountability institutions are also prone to "drift," wherein institutional functions change as a result of shifts in the external environment—such as broader changes in governance of relevant policy areas.
[36] Corbacho et al. 2016; Mungiu-Pippidi 2015; Persson, Rothstein, and Teorell 2013
[37] Berliner and Erlich 2015; Bookman and Guerrero Amparán 2009; Cejudo et al. 2014
[38] Berliner, Bagozzi, and Palmer-Rubin 2018
[39] Almanzar, Aspinwall, and Crow 2018; Berliner et al. 2021; Fox, Haight, and Palmer-Rubin 2011; Lagunes and Pocasangre 2019
[40] Levitsky and Murillo 2009
[41] Piñeiro Rodríguez et al. 2021

expertise, with evidence consistent with both compositional and behavioral mechanisms. On the supply side, we find that government obligants have increasingly responded better to more "expert" requests over time. We further find that these shifts towards expert bias are driven by precisely those obligants that receive more expert requests, indicating a feedback loop. Qualitative evidence based on interviews with grassroots and expert civil society claimants and government agents confirms the importance of both compositional changes among claimants and behavioral changes by claimants and obligants. Further, this evidence highlights the specific forms of learning, attrition, and intensification underlying these mechanisms, and the mutually reinforcing nature of these processes.

We suspect that such processes of expert evolution occur in many social-accountability institutions. The idea that these institutions privilege elite claims is not new. For instance, prior research has found that policy councils fall prey to "elite capture"[42] and that complaint mechanisms "exclude the marginalized by design."[43] In this vein, social-accountability institutions are yet another example of a space where economic elites exercise outsized influence to command greater representation, a tendency which has been well-documented in Mexican economic policy.[44] We build upon this previous scholarship by illustrating the endogenous drivers that gradually cause such a bias within institutions that were designed carefully to function for all citizens equally.

Expert-favoring evolution is not inevitable, however. Participatory policy councils and budgeting processes in Brazil, for instance, are social accountability institutions that evolved in a way to empower an ever-expanding population of ordinary citizens over two decades.[45] In the early 19th-century United States, increased petitioning activity drove legislators and officials to make new investments in responsiveness that helped make the entire institution of petitions less elite-oriented, at least for several decades.[46] In the conclusion we speculate about the conditions under which elite evolution is most likely to take hold.

II. Grassroots and Brokered Accountability Institutions

Accountability institutions have commanded attention in both scholarly and policy communities as elements of "second-generation" democratization. Most research has sought to explain either the adoption of such institutions[47] or their immediate short-term causal impacts on policy or election outcomes.[48] To the extent that scholars look at longer-term processes, they focus on the challenges of consolidation.[49] While accountability institutions are often adopted with great optimism, over time they may decline in usage or legitimacy. Accountability institutions in new democracies are often beset by "second order collective action problems" in which their

[42] Mansuri and Rao 2013, Chapter 4
[43] Hossain, Joshi, and Pande 2024, 151
[44] Palmer-Rubin 2022
[45] Baiocchi et al. 2011; Wampler 2015
[46] Carpenter 2021
[47] Berliner and Erlich 2015; Grzymala-Busse 2006; Piñeiro Rodríguez et al. 2021
[48] Banerjee et al. 2010; Dunning and others 2019; Ferraz and Finan 2008
[49] Kim et al. 2021; Mayka 2019; Relly et al. 2020

overseers—who themselves may be enmeshed in illicit activities—have little incentive to comply.[50] This problem produces a vicious cycle in which widespread corruption and the lack of sanctions are mutually reinforcing. Breaking out of such cycles typically occurs during momentous political transitions, which usher in new institutions.[51] For instance, Mexico's ATI system originated in a campaign promise by Vicente Fox, the president that interrupted that country's long-lasting one-party dominance in 2000.[52]

For institutions that rely on frequent citizen participation—as opposed to horizontal institutions, which are driven by specialized government agents—maintaining effectiveness is particularly difficult. In order for these institutions of "social accountability"[53] to contribute to a desirable equilibrium of good governance, citizens and/or organized civil society must find ongoing participation worthwhile. This continued usage is reinforced by a system that is responsive to citizen claims, for instance by providing information that they request or improving services about which they complain.[54] Conversely, breakdowns in either citizen usage or government responsiveness can route these institutions into a vicious cycle in which they ultimately become defunct.

Evolution in these institutions for better or for worse, however, need not be monolithic, but may instead occur unevenly for different classes of claims. One dimension of such uneven responsiveness—which is crucial for the principle of democratic representation—concerns the *expertise* of claimants. Expertise describes the relevant knowledge and resources to make effective use of a given institution. This may include legal know-how, policy-specific experience, computational skills, or connections to government personnel. Lack of such expertise is often a barrier to participation in many participatory settings.[55] Expertise certainly is correlated with class—wealthier citizens are more likely to have expertise, or to be able to contract the service of experts—so an expert-biased institution would exacerbate representational distortion. This malady has been amply diagnosed in the United States,[56] but is perhaps even more insidious in Latin America's oligarchic democracies.[57] This expertise advantage, however, can be tempered when ordinary citizens have expert brokers at their disposal to help navigate social accountability processes.[58]

We distinguish two ideal-typical models of accountability institutions regarding their suitability for expert vs. non-expert claimants. Particular institutions may occupy an intermediate space or exhibit variation across policy areas or administrative units.

---

50 Mungiu-Pippidi 2015; Persson, Rothstein, and Teorell 2013
51 Mayka 2019; Rothstein 2011
52 López Aceves 2013
53 Joshi and Houtzager 2012; Malena, Forster, and Singh 2004
54 Sjoberg, Mellon, and Peixoto 2017; Buntaine, Nielson, and Skaggs 2021; Dipoppa and Grossman 2020; Distelhorst and Hou 2017; Grossman, Platas, and Rodden 2018
55 Verba, Schlozman, and Brady 1995; Yang and Callahan 2007; Einstein, Palmer, and Glick 2019
56 Baumgartner and Leech 1998; Lindblom 1977
57 Cameron 2020
58 Kruks-Wisner 2022

The *grassroots accountability model* is typified by low barriers to entry, broad usage, and usefulness for lay citizens. In this model, thousands or even millions of citizens participate in the system, for instance by submitting information requests or evaluating local services. This model is desirable for its inclusiveness and capacity to generate a culture of civic engagement.[59] Participatory institutions—such as participatory budgeting in Brazil—are often designed with a grassroots model in mind and have been shown to generate a host of benefits, both in deepening democracy and in influencing policy outcomes to respond to the interest of often-marginalized communities.[60] Experimental evidence has found that interventions that encourage citizen oversight of local service delivery can lead to improvements in quality and civic attitudes.[61] The grassroots model is not designed to produce responsiveness in all contexts, however. For one, it presents challenges beyond the local scale, risking unsustainable workloads for federal government personnel due to the high volume of claims. Second, the grassroots model may break down in contentious issues of corruption or government failure, where atomized lay citizens lack the expertise to overcome government agencies' resistance.[62] Given these limitations, the grassroots model is perhaps most effective for addressing individual or local issues such as exposure to petty corruption, employment, or access to services.

The *brokered accountability model* characterizes institutions whose successful navigation requires expertise and thus whose usage is mainly restricted to specialists (journalists, elite civil society, large firms, lawyers) who act on their own account or as intermediaries for non-experts. A brokered model presents obstacles that filter out lay citizens, such as complex paperwork for filing claims, fees, or lengthy delays—i.e. "administrative burdens."[63] This model is not necessarily inoperative for non-expert claims, as has been alleged for ATI systems in the United States and Great Britain.[64] However, for such a system to be useful for non-experts, they may require the support of an expert intermediary—or "social broker"[65]—who helps navigate the institution and deploy it in collective political activity.[66] Brokerage-based institutions are thus perhaps less useful for ordinary citizens pursuing quotidian concerns, yet have the advantage of affording considerable attention and resources to address the smaller number of inputs that are successfully registered. The brokered model is thus amenable to high-profile accountability-seeking purposes, such as high-level corruption or large-scale evaluation of government programs.

The extent to which an institution tends toward one or the other of these two models can be based on formal (*de jure*) or informal (*de facto*) traits. For instance, the United States' FOIA system is formally designed for the brokered model, as users incur high financial costs and must

---

[59] Kosack, Tolmie, and Griffin 2010; Malena, Forster, and Singh 2004
[60] Baiocchi et al. 2011; Mayka and Abbott 2023; Wampler, Sugiyama, and Touchton 2019
[61] Björkman and Svensson 2009; Slough et al. 2021
[62] Lieberman, Posner, and Tsai 2014
[63] Moynihan, Herd, and Harvey 2015
[64] Kwoka 2015; Worthy 2010
[65] Kruks-Wisner 2022
[66] Fox 2015

have significant legal knowledge.[67] In contrast, Mexico's system was designed with a grassroots model in mind, prioritizing ease of use and incurring no financial cost.[68] Furthermore, Mexico's system is overseen by commissioners charged with adjudicating citizen appeals of unsatisfactory responses. However, at least as important as these formal traits are informal norms that establish how the system operates in practice. Systems that are designed for grassroots accountability may present obstacles for lay citizens in navigating the bureaucratic process or may simply be nonresponsive to inputs from citizens lacking social standing or insider knowledge. These informal norms are particularly relevant for institutions of accountability, which are subject to sizeable variability in their in-practice operation as they grant a high level of discretion to authorities.

To study whether social-accountability institutions can evolve over time, to become closer to one model or the other, we focus on the case of Mexico's ATI system. This case selection is driven by the opportunity to study a social-accountability institution both over a long period of time (2003-2019) and with fine-grained and detail-rich data on individual citizen-government interactions. Our expectation is that Mexico's ATI system has become increasingly responsive to expert claims over time. This expectation is based on anecdotal reporting by activists who engaged with this system over its two-decade lifespan. Such an evolution would resemble the phenomenon of "elite capture," which has been well-documented in research on some participatory institutions.[69] At our research project's onset, we had strong reason to think that Mexico's ATI system had evolved toward a brokered institution, but we remained open to the opposite finding. In fact, as we describe below, officials at Mexico's ATI institution have devoted substantial effort to promoting broad participation by ordinary citizens. The three hypotheses below allow us to test our expectation of evolution towards an expert-biased institution.

Hypothesized Drivers of Evolution

To explain this gradual process, we draw on historical institutionalist approaches outlining mechanisms that alter the *de facto* functioning of the system independent of any change in the formal institutional framework.[70] Contingent, gradual, and interactive, such processes may unfold quite differently across government agencies, depending on the iterated actions and reactions of citizen claimants and government obligants. Given the high degree of discretion in enforcement, gradual change in social-accountability institutions mainly operates through *conversion*, a process wherein "existing institutions are redirected to new purposes, driving changes in the role they perform and/or the functions they serve"[71] or through *drift*, in which "changes in the operation or effect of policies occur without significant changes in those policies' structure."[72]

---

[67] Kwoka 2015
[68] Bookman and Guerrero Amparán 2009; Michener 2013
[69] Mansuri and Rao 2013, Chapter 4
[70] Mahoney and Thelen 2010; Pierson 2011; Thelen 2003
[71] Thelen 2003
[72] Hacker 2004, 246

If indeed Mexico's ATI system has evolved over time to become closer to a brokered system, then we should be able to observe empirical evidence of three different drivers of system-level change. We theorize these drivers here, developing specific hypotheses.

We first predict *demand-side* drivers, having to do with the average level of expertise among citizen claimants. We address two subsidiary mechanisms. Through repeated experiences with the institution, claimants may undergo *behavioral* changes. This is particularly relevant for expert claimants who, for instance, learn how to most effectively word an information request, such as by citing relevant legislation. An additional mechanism relevant to claimants (but not to obligants in the present context) is *compositional* change, by which certain groups of claimants come to constitute a greater or lesser portion of total claims. This change may occur on an extensive margin, as when positive experiences spread through word-of-mouth leading to greater use within a given sector, or when some types of claimants "drop out" after poor experiences. But compositional changes may also occur on an intensive margin, as when certain individual claimants increase (decrease) their number of claims in response to prior success (failure). We may expect higher levels of attrition among non-experts, who lack the time and knowledge to confront administrative hurdles.

H1 (Demand): Demands exhibit gradual change in the level of expertise and will increasingly be characterized by higher levels of expertise.

We additionally analyze *supply-side* drivers of evolution in which government obligants alter their behavior in ways that affect patterns of responsiveness. Such behavioral change may consist of capacity building, in which obligants accrue resources or develop processes to more effectively handle citizen claims. In the ATI realm, capacity building entails the development of systems and hiring or training of staff for managing records and large numbers of information requests. Behavioral change may also be strategic, wherein obligants prioritize which claims to be more or less responsive to. Such choices respond to incentives such as the likelihood of appeals for poor responses, sanctions for noncompliance, or professional reprisals for aiding corruption investigations into superiors.[73] We may expect supply-side change to favor expert claims, who are more likely to appeal unsatisfactory responses.

H2 (Supply): Responsiveness to expert claims improves over time relative to responsiveness to non-expert claims.

These drivers of change are possible in any institution featuring repeated interaction between citizens (or non-state actors) and the state. We may expect that evolution toward experts (or economic elites) is the norm in many cases in which the adoption of new state institutions is followed by the proliferation and maturation of elite civil society or corporate actors that specialize in navigating that institution. Tarrow, for instance describes the "coral reef"[74] of international human rights NGOs that were founded to navigate the international organizational apparatus in the late-twentieth century. However, many of our implicit theories of state-society relations assume a grassroots tendency in new institutions, as word of mouth spreads and

[73] Berliner et al. 2021; Kim et al. 2021
[74] Tarrow 2005

ordinary citizens develop aptitudes. Participatory budgeting in Brazil,[75] and the rise of petitioning in the early 18th-century United States,[76] are two examples of new institutions that successfully spurred capacity building and contagion to include increasingly broad participation. Demonstrating a different form of compositional change, a recent analysis of Wikipedia editors shows how gradual "population loss" over time in response to differential success rates in content disputes led to the eventual predominance of objective—as opposed to fringe—editors.[77]

Feedback Effects

Finally, these supply and demand-side drivers of change interact, producing "feedback effects"[78] that accelerate tendencies toward either a grassroots or brokered model of responsiveness. For instance, compositional change among claimants may influence obligants' choices about resource investments. Thus, as non-expert claims come to predominate, obligants may invest in staff and procedures to handle a large number of simple claims. Conversely, if expert claims expand, obligants may invest in resources to handle more complex types of claims, such as more specialized personnel and digital resources. As obligants become increasingly adept at handling certain types of claims, their ability to respond to other claims may also atrophy, spurring the disfavored class of claimants to further withdraw.

Obligants' responses are conditioned by policies that create incentives (rewards and sanctions) for effective responsiveness. The above-described scenarios assume that the institution offers such incentives. However, obligant learning need not only function to improve responsiveness; it may function to avoid responding to certain types of claims, such as those most likely to be damaging to agency leadership. Such a response by obligants could discourage these types of claims.

Entering or exiting the system (compositional change) is not the only response available to claimants. They may also adapt their strategies of claim making in response to prior experience. After repeated attempts at claim making, claimants may find that certain types of issues receive better responses from obligants, or that their claims are more successful when presented in a certain way. For example, if obligants prove most responsive to expert claims, some claimants may learn to include markers of expertise in their claims, such as legal language or insider information. Such behavior would accelerate tendencies toward a brokered accountability system even in the absence of compositional change. Over time, these processes of endogenous change can accumulate, fundamentally shifting the functioning of the institution in terms of whom in society it best serves.

H3 (Feedback between demand- and supply-side changes): Agencies experiencing greater increases in the level of expertise of claims they receive will exhibit greater shifts towards responsiveness in favor of more expert claims.

---

75 Baiocchi et al. 2011
76 Carpenter 2021
77 Steinsson 2024
78 Pierson 1993

We test these hypotheses using quantitative and qualitative evidence to track changing patterns among both claimants and obligants over time. Quantitatively, we examine shifts in the volume and expertise of information requests, differential responsiveness to these requests, and interactions between these. Qualitatively, we detail the distinct experiences and adaptations of expert and non-expert requesters, as well as changing behavior of government officials.

III. Mexico's Access-to-Information Institutions

ATI requests to the Mexican government constitute a particularly opportune venue to observe over-time change in social-accountability processes. These requests offer a source of massive and highly granular data about citizen-government interaction. Over nearly two decades, Mexican citizens have made an average of nearly 200 information requests *per day* to the federal government. More than twenty ministries and agencies regularly receive over 1,000 requests per year, offering a wide range of policy areas in which to observe changes in responsiveness. Mexico's ATI system was founded with a vision of establishing the right to information as a central component in Mexican democratization, following the 2000 demise of one-party rule of the federal government.[79] The Federal Law on Transparency and Access to Public Information was passed in June 2002, taking effect one year later. The law has often been praised as one of the strongest in the world, in particular for its independent information commission, easily accessible online platform, high usage volume, and quick response times and low denial rates.[80] The *Instituto Federal de Acceso a la Información* (IFAI; later renamed INAI) was created to promote awareness and usage of the new law, to monitor bureaucratic compliance, and to hear appeals. The law also created an online information system, unique in the world at the time, for citizens to file requests and receive responses.[81]

In Mexico's federal ATI system, the obligants are the federal government agencies that receive information requests.[82] Government agencies established Transparency Units, which developed diverse systems for handling large numbers of requests.[83] Agencies are officially evaluated and compared by IFAI on the basis of "success rates"—how often they respond to requests positively, reporting that they have provided the information requested.[84] This system incentivizes Transparency Units to prioritize efficiency, responding to as many requests as possible. However, the large number of information requests precludes IFAI from exercising

---

[79] Cejudo et al. 2014
[80] Bookman and Guerrero Amparán 2009; Michener 2011
[81] This system has been redesigned and renamed over time: Sistema de Solicitudes de Información (SISI, 2003-2008); INFOMEX (2008-2017); Plataforma Nacional de Transparencia (2017-present). We use the term INFOMEX throughout. Even non-electronic requests are managed through the INFOMEX system, with agency officials entering the relevant information.
[82] Our study only investigates federal government agencies. State governments adopted similar ATI systems in the early 2000s (Berliner and Erlich 2015). Although they were later brought into the INFOMEX system, we focus on federal requests in order to study a similar set of obligants over time. Mexico's 2016 transparency reform also made labor unions and political parties subject to transparency requirements, but we similarly exclude these from analysis.
[83] Ríos Cázares, Castañeda, and García 2017
[84] Fox, Haight, and Palmer-Rubin 2011

individualized oversight. As a result, agency personnel exercise significant discretion in handling individual claims. Officials utilize this discretion in balancing conflicting incentives for and against responsiveness, and often withhold information that could be damaging to agency leadership.[85] However, Transparency Unit personnel also value projecting an image of competence and many are personally committed to transparency norms.[86]

IFAI's designers envisioned an institution that would be accessible to expert and non-expert claimants alike. High-profile civil society and academic figures had been invested in the system from the start, motivated by the potential of ATI to enable corruption investigations.[87] Thus, the main challenge was to expand knowledge and use of the system among ordinary citizens amidst a culture wherein public servants were understood to be "owners of government information."[88] The hope was that this user-friendly institution would empower ordinary citizens as subjects in Mexico's newly pluralistic democracy. For instance, the original transparency law stipulated that indigenous citizens could make information requests in their native language and would have access to interpreters to help understand the responses.[89]

A fledgling IFAI also encouraged grassroots organizations from throughout the country to make public information requests a central part of their advocacy activities. Proyecto Comunidades was an initiative by which IFAI hired facilitators to partner with low-resource organizations and train them to make information requests and use public information in their advocacy. In an interview, the director of Proyecto Comunidades attested to the social empowerment afforded grassroots communities in the early years of IFAI:

"The project lasted from 2005 to 2009. And what it showed was that marginalized communities only needed a process of accompaniment to be able to exercise their right to access information...For us this is cause for enormous pride, because it shows not only their interest in knowing, but also that they saw in those stacks of papers a source of power that they did not have before—the power to confront the authorities without being misled."[90]

A 2006 international evaluation of IFAI highlighted its orientation to broad use by Mexican citizens and organized civil society:

"There is an emphasis on extending the work of IFAI to as many groups and citizens as possible to make the Transparency Law effective...The Transparency Law is largely an accomplishment of the Mexican civil society that lobbied and won its passage. It is a law that stands for unprecedented public openness in Mexico's history. The underlying hopes are that each citizen

---

85 Berliner et al. 2021; Gill and Hughes 2005
86 Cejudo et al. 2014, ch.3
87 Lachenel and Ruiz Guerra 2013
88 Merino 2013
89 López Aceves 2013, 11–14
90 Author interview with Maylí Sepúlveda, conducted online, December 10, 2020. Translation by Author.

throughout Mexico's thirty-one states benefit, directly or indirectly, from the functioning of the transparency laws and the discipline and educative support of IFAI."[91]

In sum, Mexico's ATI system was founded as a multi-purpose institution, with characteristics of both a brokered and a grassroots accountability institution. There are certainly signs that the efforts to expand usage among ordinary citizens have been successful. According to a recent government survey, roughly a quarter of citizens have accessed information through federal government agencies' transparency portals.[92] 72 percent of respondents to this same survey agreed with the notion that all Mexicans have a right to the information generated by the government. The INAI (as IFAI was subsequently renamed) itself reinforces this message and in 2019 launched a program called the National Plan for Socializing the Right to Information (PlanDAI), including a mass media campaign and training sessions for grassroots civil society actors.[93] Evidence also exists, however, that Mexico's ATI system has professionalized, increasingly becoming the purview of experts. In the early 2000s, an expansive ecosystem of CSOs and academics emerged, joining together in the Network for Accountability (Red por la Rendición de Cuentas).[94] As detailed below, these organizations have developed sophisticated expertise in advocacy campaigns using information gathered through ATI requests. Furthermore, levying information requests has become an important tool for private firms that do business with the government or are subject to regulation.[95]

IV. Data and Measures

We use data from the publicly available system maintained by Mexico's information commission, including every request for government information filed with Mexican federal government entities since the 2003 implementation of Mexico's ATI law. Although the vast majority (over 97%) of requests are filed electronically via this free online system, manual requests filed on paper are also entered into the system by officials. We use requests filed through the end of 2019 to avoid complications from the COVID-19 pandemic beginning in early 2020. Additionally, we exclude requests filed with a set of non-federal entities (such as state governments, the Supreme Court, and political parties) which were only incorporated into the INFOMEX system following a 2015 policy change. The first panel of Figure 1 shows the growing overall annual volume of requests for government information filed with federal entities from 2003 through 2019.

FIGURE 1 HERE

These data include, for each request, the name of the government entity to which it was directed, the date of request, the full text of each request including any attached request files (which we scraped, digitized, and merged into the request text), the official category of response, and text of response including attached files. The data also include the municipality and postcode of the

---

91 Sobel et al. 2006, 2

92 INEGI 2020

93 PlanDAI 2020

94 See: https://www.rendiciondecuentas.org.mx/que-es-la-red-por-la-rendicion-de-cuentas/

95 Lachenel and Ruiz Guerra 2013

requester. Initially mandatory, a 2016 policy change allowed requesters to omit this geographical information; 34% of requests left these location fields blank from 2017-2019. As such, in some analyses using location-based data we assess only requests through 2016. The publicly available data do *not* include individual requester IDs, preventing us from knowing which specific requests were filed by the same individual or organization.

Measure of Request Expertise

Testing our hypotheses requires additional measures: the level of expertise exhibited in the text of each request and responsiveness of government obligants to each request. While ideally we would measure request expertise using the actual identity of each user, this information is protected by INAI. Further, given the INFOMEX system's provisions to protect requester identities, responding officials are typically only able to infer requester characteristics on the basis of the request texts themselves, making our text-based expertise measure an appropriate proxy. We worked with a team of six Mexico City-based research assistants, each with professional experience with Mexico's ATI system, to hand-code a random sample of nearly 5,000 requests for a variety of characteristics. We then used multi-label prediction[96]—a machine learning method for the supervised learning of multiple correlated and non-exclusive labels—to produce predicted probabilities of these characteristics across the full set of requests. See Appendices K and L for additional information.

To produce a text-based measure of request expertise, we combined nine of these predicted measures. Each of the hand-coded requests was coded for the presence of *formal*, *legal*, or *technical* language, yielding three separate predicted language measures. Requests were also coded for any mention of specific *persons*, *places*, *dates*, *institutions*, *organizations*, or *documents*; yielding six predicted specificity measures. Our measure of request expertise is thus based on the *revealed* expertise suggested by the request text itself, on the basis of using more formal, technical, or legal language; or greater specificity in describing the requested information. Low expertise, on the other hand, is signalled by the use of more ordinary language and description of the information sought in more general terms.

Our Expert Request Score is the of sum these nine measures, after first dividing the sum of the language measures by three and the sum of the specificity measures by six, such that the two sets of measures are weighted equally in the ultimate score. As an alternative, in robustness checks we instead use the primary component (accounting for 30.5% of total variance) of a principal-components analysis applied to all nine measures.

As a check on the validity of this measure, in Appendix B we graphically compare it with the economic marginality (an index produced by Mexico's Consejo Nacional de Población) of the municipalities from which requests are filed. As expected, we see that less-expert requests are filed from poorer municipalities, and more-expert requests are filed from wealthier municipalities.

---

96 Erlich, Dantas, et al. 2021

Measures of Responsiveness

We construct five request-level measures to capture the different practical and legal dimensions of responsiveness in the Mexican ATI context. The first two of these use official response designations, consistent with other scholarly work and policy evaluations of Mexico's ATI system.[97] First, "Days to Response" measures the logged number of working days (corrected to account for weekends and national holidays) from request to response. Not only do slow responses violate legal time limits, but they also render the response less useful to claimants who may have time-sensitive reasons to seek information. Second, "Official Inexistencia" measures the occurrence of a frequently misused official response designation,[98] claiming that the requested information does not exist.

Three additional measures employ machine predictions of the *quality* of delivered information, drawing on the hand-coding procedure described above. Our third measure, "True Inexistencia," augments the Official Inexistencia measure to include responses that were misleadingly classified as having provided the information. This variable takes values of one where *either* the official response designation, or the actual text of the response letter, claims that the information does not exist. The phenomenon of "misleading inexistencia," whereby responses that are formally categorized as having delivered information in reality only say that no information exists, is much maligned among policy and advocacy communities in Mexico.[99] Fourth, "Uninformative Response," takes a value of one where the request received either an official denial or a response that is officially designated as positive but in fact provides less than half of the information requested (using predictions based on hand-coding of the information contained in the response). This variable thus gauges whether the requester received the information that they sought, even if the response was legally compliant. Finally, our fifth measure, "Noncompliant Response," is only meaningfully observed for responses that are officially designated as positive. Among these, this measure takes a value of one where no information is provided, meaning that there was clear legally non-compliant official behavior in designating the response as positive.

Control Variables

In some analyses we use control variables derived from the predictions of additional hand-coded request characteristics. These include multiple indicators of the appropriateness of the request, the complexity of the request, and the theme of information requested. We also control for mechanical characteristics of requests, including the logged word count, the ratio of characters to word, whether the request was filed electronically or manually, and did or did not include an attachment. More information is available on all of these in Appendix A, along with summary statistics of all variables in Appendix E.

---

[97] Almanzar, Aspinwall, and Crow 2018; Lagunes and Pocasangre 2019; Cejudo, Ríos, and López-Ayllón 2010
[98] Fox, Haight, and Palmer-Rubin 2011
[99] Fox, Haight, and Palmer-Rubin 2011

Finally, in some analyses we also measure the wealth of the municipalities from which requests were filed (bearing in mind that after 2016 some requesters left this field blank), by merging requests with the municipal economic marginality index produced by Mexico's Consejo Nacional de Población. We use marginality figures from 2000 to ensure that these precede the period under study and that differences reflect only cross-sectional variation.

V. Analyses and Results

Assessing over-time change among claimants and obligants require distinct empirical analyses. We first focus on understanding shifts over time in the expertise of claimants, and then turn to focusing on changes in obligant responsiveness.

Assessing Claimant Compositional and Behavioral Changes

Our first analysis pertains to hypothesis one—evaluating over-time change in the expertise level of information requests. The second panel of Figure 1 shows the average request expertise score over time, indicating a clear trend towards greater expertise, in support of hypothesis one. This trend is substantial in magnitude: the difference between the average score in 2019 and the average score in 2003 is 0.33 of a standard deviation of the expertise score across all individual requests.[100] Further, the median request in 2019 has a higher expertise score than 75.5% of all individual requests in 2003. This shift occurs gradually and does not align with external political or formal policy changes. Increasing request expertise can be driven by two potential mechanisms: (1) compositional change, whereby more elite users make more intensive use of ATI requests, while more grassroots users make less intensive use or even give up entirely; or (2) behavioral change, whereby some claimants increasingly incorporate markers of expertise.

Models in Table 1 help evaluate these two mechanisms. Importantly, we do not seek to formally test between them as we think it most likely that both are at work. We do, however, seek to demonstrate empirically that it is plausible that a learning mechanism of behavioral change accompanies a mechanism of compositional change. The first model shows the basic finding that requests exhibit increasing expertise over time. However, this alone does not indicate learning, as this pattern could be driven by compositional changes of less expert users dropping out while more expert users differentially enter, remain, and/or file more requests. If we had individual user IDs, an ideal analysis would include fixed effects for every single user and assess whether the within-user coefficient on Year remained positive. But since user IDs are protected by INAI for personal data protection reasons, we must take an alternative approach.

TABLE 1 HERE

Instead, to discern the degree to which increasing expertise is attributable to claimant learning, we add fixed effects for a series of increasingly fine-grained categories that mimic individual users by capturing specific use-types. Model 2 adds fixed effects for each municipality, while Model 3 adds fixed effects for each postcode (treated as nested within municipalities). Model 4

---

[100] Figure C.2 in Appendix C presents histograms of the relative distributions of expertise score in 2003 and 2019.

uses fixed effects for each of 60,188 agency-municipality combinations, while Model 5 uses fixed effects for each of 287,811 agency-postcode combinations. Finally, Models 7 and 8 use fixed effects for each of 122,309 agency-municipality-theme combinations and for each of 436,251 agency-postcode-theme combinations, respectively. If we could assume that a single user only ever files requests on a single theme to a single agency, and that no more than one such user lives in each postcode, then this analysis would be identical to the ideal analysis using fixed effects for individual user IDs. Of course, that is not the case in reality, and we do not claim that this assumption is valid. As such, our results may still capture some compositional changes over time within these categories.

Nonetheless, results are highly consistent across all these models. In each, the coefficient on Year remains positive and statistically significant, indicating that requests increase in expertise over time, within each of these categories as well as across them. Notably, the magnitude of the coefficient generally declines as the categories become more specific (moving closer towards the ideal situation of fixed effects for each user). However, even the smallest coefficient (0.0028) is still 80% of the size of the original coefficient in Model 1 with no fixed effects. This suggests that while compositional changes do account for a share of the overall increase in request expertise over time, a substantial share is also accounted for by behavioral changes. These coefficients are also substantively large in size. Even in the most fine-grained model, multiplying the coefficient by sixteen years suggests a within-category over-time shift in average expertise equivalent to 38.7% of a standard deviation of the request expertise score (0.12).

Assessing Obligant Behavioral Changes

We now turn to the second hypothesis, evaluating behavioral changes in the obligants of Mexico's ATI system, federal government agencies. We seek to assess both the extent to which responsiveness to different types of requests changes over time, and whether or not these changes reflect a mechanism of learning. Compositional change is not relevant here given the non-voluntary nature of this policy for Mexican government agencies.

We model measures of responsiveness on request expertise. Given that more expert requests are typically longer, more complex, and more likely to include multiple queries than the average lower-expertise request, we control for these factors and test whether any differential improvements over time in responsiveness to more and less expert requests are indeed statistically distinguishable. We further assess if the extent of these trends are attributable to agency-specific exposure to receiving more or less expert requests—which would be consistent with learning-driven feedback loops by officials.

All models of responsiveness include fixed effects for each agency, as well as a series of controls that we detail in Appendix A. For the logged days-to-response outcome, we use a linear model, and for all other dichotomous outcomes we use linear probability models given the large number of fixed effects. To ensure the results are not unduly shaped by entities that receive relatively few information requests (such as individual government-run clinics and universities), we restrict the sample to only entities that appeared in the top 100 by request volume in at least one year of the 2003-2019 period. The resulting sample includes requests to 209 agencies, which receive on average 544 requests per year, with a range of 38 to 7,468 requests. In a robustness check in

Appendix I, we instead use the full set of all 818 agencies that appear in our data, even though many of these are relatively small entities.

In Table 2, we show results for our first five models of request-level responsiveness, with results mainly in support of hypothesis two, pertaining to evolution in obligant responsiveness. Each model uses a different measure of responsiveness as the dependent variable: logged working-days-to-response, official "inexistencia" responses, true "inexistencia" responses, uninformative responses, and noncompliant responses. The latter three use information from the machine-learning-generated predictions of hand-coded response characteristics. For all outcomes, higher values indicate poorer responsiveness. Across all five models, the base coefficient for Expert Request Score is positive, and statistically significant in four of these five. This suggests that, initially, more expert requests received poorer responses than less expert requests. However, in all five models, the interaction term between Expert Request Score and Year is negative (and statistically significant in four of these five). This indicates that, over time, the responses received by more expert requests improved faster than those received by less expert requests.

TABLE 2 HERE

The first two panels in Figure 2 illustrate the substantive effects of these temporal change in responsiveness interacted with expertise, based on Models 1 and 3 in Table 1. We compare otherwise-equal but less-expert requests—with the expertise score set to one standard deviation below its mean—and otherwise-equal but more-expert requests—with the expertise score set to one standard deviation above its mean. To clarify, these are only hypothetical requests (holding all other variables at their means) in order to illustrate the results, whereas actual requests reflect a wide range of expertise scores. In the first year of the system, information requests with low expert scores were predicted to receive responses in 10.8 days, while requests with high expert scores were predicted to receive responses in 14.2 days. Yet after ten years, expertise appeared not to be associated with response time at all. And by 2019, expert requests were receiving responses 2.1 days sooner on average than non-expert requests. In all, over this period agencies became 2.6 days faster at responding to expert requests, yet 2.8 days *slower* at responding to non-expert requests, holding other factors constant.

The second panel in Figure 2 shows similar patterns for "inexistencia" denials that claim the requested information does not exist – whether or not this was officially recorded or misleading. This form of unresponsiveness grew worse for both expert and non-expert requests over time, yet this trend was steeper for non-expert requests. While initially non-expert requests were slightly favored in receiving fewer such responses, by 2019 non-expert requests were 4.3 percentage points more likely to receive this form of response than were expert requests. Appendix G shows similar plots for models with the remaining outcome variables, and Appendix H instead presents the same results using interflex plots as suggested by.[101]

FIGURE 2 HERE

Assessing Feedback Effects

---

[101] Hainmueller, Mummolo, and Xu 2019

Finally, we test our third hypothesis, pertaining to the interaction between request expertise and responsiveness bias. If the trends over time are indeed shaped by the feedback loop between changing requester composition and behavior on the one hand, and changing obligant behavior on the other, then we should see greater changes in response behavior in those agencies for which request characteristics have themselves shifted more. In this analysis, we replace the Year variable with a new variable measuring, for each agency, the average expert score across all requests received in the *previous* year. We then interact this measure with the same request-level expertise score. This interaction helps capture potential positive feedback loops whereby changing request characteristics in turn drive behavioral change on the part of responding officials.

In support of hypotheses three, four of the five interaction terms in Table 3 are negative and statistically significant, indicating that agencies demonstrate particularly large improvements in responsiveness to expert requests when these requests constituted a large share of their total requests in the prior year (reflected in a higher average expert score). This finding suggests a learning mechanism whereby agencies facing larger volumes of more expert requests are more likely to undertake behavioral changes which differentially improve responsiveness to such requests. As detailed with qualitative evidence below, these behavioral changes include: investing resources or staff in Transparency Units, improving internal systems for records management and request tracking, and discerning requesters that are more likely to appeal unsatisfactory denials.

TABLE 3 HERE

The bottom two panels in Figure 2 present substantive effects based on the results of Models 1 and 3 in Table 3. These illustrate the feedback effects between the composition of agencies' requests and patterns of differential responsiveness to more and less expert requests. Where agencies tend to receive less expert requests, their responsiveness shows no biases or even tilts in favor of less expert requests. But where agencies tend to receive more expert requests, their responsiveness shifts in favor of those more expert requests. At its highest points, this can result in biases in favor of more expert requests as great as 5.4 days in response times or 14 percentage points in the proportion of "inexistencia" responses.

In Appendix I, we show several robustness checks for both sets of models. First, we use an alternative measure of request expertise based on a principal-components analysis of the same underlying language and specificity measures. Second, we use the full sample of agencies rather than omitting those smaller entities for which our theoretical framework is less applicable. Third, we include additional indicators for time-specific periods reflecting the changing presidency and the same legal and constitutional changes shown in Figure 1. Results are highly similar in all cases. Additionally, following Fong and Tyler,[102] we repeat our main models using the original hand-coded versions of each variable rather than any machine-learning predictions (with the exception of the Agency Lagged Average Expertise measure, which we do not recalculate due to the sparsity of coverage of hand-coded observations across individual agency-years). Results on

---

[102] Fong and Tyler 2021

our main coefficients of interest remain consistent in sign, but in some models are no longer statistically significant due to the much smaller sample size. Finally, we also conduct interaction term tests using the Hainmueller, Mummolo, and Xu method.[103] These show that although the interaction terms depart from perfect linearity, all remain monotonic for each model where the interaction term itself is statistically significant.

In Appendix F, we also return to the demand-side results presented in Table 1, to test for feedback effects in the reverse direction: of agency responsiveness on request expertise. Analogously to our tests for feedback effects thus far, we replace the time indicator that appeared in the models in Table 1 with measures of the average responsiveness corresponding to requests filed from each municipality in the prior year. We repeat this for each of the five different response measures we use. We find clear evidence of expertise-increasing feedback effects for three of these responses. Requests filed from municipalities where, in the previous year, responses tended to be more delayed or more likely to receive "inexistencia" denials, tend in turn to have significantly higher expertise scores. We find no significant patterns for noncompliant responses. Finally, we see effects in the reverse direction for uninformative responses. This even may help explain why results for this form of responsiveness diverged from the others in most of our results in this paper. Receiving an only partially informative response may not induce lay requesters to either give up or change behaviors in the same ways that appear to be driven by receiving delayed or denied responses, but may instead instigate more follow-up requests of similar expertise levels.

Overall, these results demonstrate the importance of our theoretical framework emphasizing informal functional changes over time in social-accountability institutions. We find clear evidence for all three hypotheses pertaining to evolution in (1) demand expertise, (2) responsiveness bias, and (3) and interaction between these two. All three of these processes have pushed Mexico's ATI institutions in the direction of a brokered-accountability model, wherein more claims exhibiting higher levels of expertise both make up increasing proportions of total claims and receive increasingly superior responses vis-à-vis less-expert claim. Our results also offer evidence consistent with mechanisms of learning by both claimants and obligants, and—to a lesser extent—of change in the composition of claimants.

In the next section, we turn to qualitative evidence to both confirm these results and to further investigate the specific mechanisms involved in these shifts.

VI. Qualitative Evidence

In this section, we summarize field research evidence to illustrate the causal mechanisms underlying the shift to a brokered-accountability model. We draw on semi-structured interviews with representatives of 14 Mexican civil society organizations (CSOs) and staff employed in federal government agencies' Transparency Units. Interviews with CSOs were conducted over Zoom by two of the authors between September 2020 and January 2021. Organizations were selected to represent variation in sector (education, environment, gender rights, health, human rights). Within each sector, we interviewed at least one "elite" organization—with a high level of

[103] Hainmueller, Mummolo, and Xu 2019

human and financial resources—and one "non-elite" organization with lower resources.[104] This evidence most strongly supports the mechanism of *behavioral changes (learning)* by both claimants and obligants.

Obligant Behavioral Change: Favoring Expert Users

Throughout the first decade of the Mexico's ATI institutions, federal government agencies developed internal systems for handing information requests and devoted resources to build capacity in information management.[105] Interviewed transparency bureaucrats noted that during the presidency of Felipe Calderón (2007-2012)—Fox's successor—a culture of transparency began to develop in government agencies. Transparency Unit staff members stated: "with institutions like INAI, we have become more committed to transparency"[106] and "since these transparency laws started to come into effect now everybody knows I have to do everything correctly."[107]

As the ATI system became more sophisticated, government agents developed two main tendencies favoring expert requests. First, obligant staff became increasingly attuned to the imperative to avoid disclosing potentially damning evidence of corruption or governance failure following high profile scandals that were uncovered by investigative reporters and civil society organizations using information obtained from ATI requests. For instance, one Transparency Unit director admitted that their superiors ordered them to deny any information request related to Brazilian construction company Odebrecht—even those for information that was clearly public—because such information could reveal evidence of kickbacks to agency leadership or elected leaders.[108] Additionally, agencies became increasingly adept at detecting the identities of claimants, either by violating the confidentiality of the request or by making inferences based on the language used in the request text. Transparency Unit staff reported that they grew able to predict which information requests would likely be followed by appeals to the INAI in case of an unsatisfactory response: those that revealed expertise or insider knowledge on the part of the claimant.[109]

The second mode of obligant learning corresponded to strategies of *covert non-responsiveness*—avoiding responding to information requests while appearing to comply with the law. Interviews revealed a host of such strategies, including: (1) requiring claimants to travel to agency premises in person to pick up hard copy documents;[110] (2) requesting additional information to fulfill the request; (3) providing information completely different from that requested; (4) claiming that the

---

[104] See Appendix J for further details on case selection, interview methodology and ethical considerations. We do not provide any identifying information about Transparency Unit staff because these participants were promised anonymity.
[105] Author interviews with TU1, Mexico City, March 14, 2017; and TU4, Mexico City, March 13, 2017.
[106] Author interview with TU2, Mexico City, March 9, 2017. Translation by Author.
[107] Author interview with TU3, Mexico City, March 9, 2017. Translation by Author.
[108] Interviews, TU2, TU3.
[109] Interview, TU1.
[110] Interview, TU2.

request corresponds to a different agency; (5) providing an enormous amount of documents, forcing the claimant to search for the information requested; (6) providing information in illegible or inconvenient formats, such as scanned images of spreadsheets rather than manipulable Excel files. These tactics further disfavor non-expert claimants, who lack the time, resources, and know-how to understand whether the response was legitimate. Non-experts also face obstacles to legally appeal the responses through *recursos de revisión*, which must be filed within 15 days of the official response and provide a legal justification for why the response was inadequate.[111]

The most frequently cited tactic of covert non-responsiveness was to claim that the information requested does not exist at all. This strategy is particularly effective because it is extremely difficult for the claimant to prove in an appeal that the information exists. Every single civil society representative interviewed expressed frustration with these "inexistencia" denials. As one activist put it, "I got the impression that they kept giving us that response because we couldn't find a way to get around it."[112] Beginning in 2010, agencies began labeling a higher percentage of the "inexistencia" denials as having provided the requested information, after a controversial decision by INAI commissioners authorized such responses in cases where the agency was not legally required to have the information.[113] This decision allowed agencies to deliver "inexistencia" denials without detection by INAI overseers. Following this, the total share of such denials grew from 12.6 percent (of all requests) in 2010 to 20.3 percent in 2019, even as the officially reported rate actually decreased from 7.2 percent to 4.3 percent.

Non-Expert Claimants: Attrition with Replacement

Given shortages in resources and expertise, non-expert claimants are limited in adapting to these obfuscating tactics by obligants. Where grassroots (i.e. non-expert) organizations have had enduring success in use of the ATI system, it has typically been because they partnered with much more experienced organizations that were capable of mobilizing resources, contacts and expertise.[114] With these exceptions, grassroots civil society organizations were more apt to simply drop out of the system after repeated failed requests, despite ongoing information needs. A staff member at one organization explained that to make effective use of the system, "you'd have to learn all of the reasons that they can deny your request and learn how to contest them. Honestly, who has time for that?"[115] Such reports of "throwing in the towel" suggest that compositional change is potentially part of the explanation for the increasingly expert-serving nature of Mexico's ATI system. A reduction in requests by non-expert users may be further accelerated by the increasing availability of some of the most-commonly requested types of information on agency websites,[116] obviating the need to submit requests. However, the ever-

---

[111] Author interview with Mejora tu Escuela, conducted online, December 8, 2020.
[112] Interview, Mejora tu Escuela, Translation by Author.
[113] This decision is codified in Criterio 07/17. See:
https://www.ieem.org.mx/transparencia2/pdf/fraccionI/criterios/07-17.pdf
[114] Author interviews with Controla Tu Gobierno, conducted online, November 13, 2020; DVVIMSS, conducted online, November 13, 2020; Maylí Sepúlveda.
[115] Author interview with Poder, conducted online, November 25, 2020. Translation by Author.
[116] Cejudo, Ríos, and López-Ayllón 2010

expanding population of users suggests that non-expert claimants continue to constitute a large portion of users of Mexico's ATI system.

Expert Claimants Develop Best Practices

The most important mechanism favoring the brokered model on the claimant side is learning (behavioral change) by expert claimants. Interviewed expert civil society organizations reported several "best practices" that they developed through trial-and-error. First, these experts capitalize on inner knowledge of agencies, facilitating requests that are specific (e.g. by mentioning documents or officials by name) and appropriate to the agency in question. Second, experts signal expertise by using technical or legal language. Representatives of expert organizations learned that when they applied these strategies, their information requests were more likely to be successful:

"One of the first lessons we learned was the need to be very specific in what we were requesting."[117]

"Our great advantage is that we have all the context behind the question that we are asking, making it easy to phrase the perfect question in a way that they cannot evade responding."[118]

"I have seen that when your request has a legal justification, where you expressly reference the article and section of the law, you get a more favorable response. The authorities take your request more seriously."[119]

Finally, expert claimants describe capitalizing on superior human resources and political connections. Several representatives reported that contacting agency personnel directly by phone or email to follow up on information requests was the only way that they were able to finally access the requested information.[120] An increasing proportion of expert requests is also attributable to an intensive-margin compositional change as certain expert "super requesters" developed "information request factories," areas dedicated to submitting and following up ATI requests, sometimes as many as thousands per year.[121]

To conclude, qualitative evidence affirms the evolution of Mexico's ATI system to a brokered-accountability institution and highlights the importance of claimant and obligant learning—alongside change in the composition of claimants—as the principal drivers of this evolution. Expert claimants—such as highly resourced civil society organizations based in Mexico City—codified a set of "best practices" to navigate the system. At the same time, agency personnel began responding more effectively to requests that signaled expertise and thus a high likelihood of appealing unsatisfactory responses. Obligants also devised a set of tactics for "covert non-responsiveness" that were particularly burdensome to non-expert claimants. To some extent,

---

[117] Interview, Mejora Tu Escuela, Translation by Author.
[118] Author interview with Nosotrxs, conducted online, November 6, 2020. Translation by Author.
[119] Author interview with Oceana, conducted online, December 14, 2020. Translation by Author.
[120] Author interviews with Equis Justicia, conducted online, November 12, 2020; Nosotrxs.
[121] Author interviews with Controla tu Gobierno, Equis Justicia, Mejora Tu Escuela.

these behavioral changes were accompanied by changes in the composition of claimants, particularly attrition by non-expert claimants who got fed up with the system.

VII. Conclusion

This paper analyzes the gradual internal processes of change that affect participatory institutions. While much interest remains in the question of how these institutions consolidate and persist over time, our focus is different. We address endogenous change that affects bias in favor of certain classes of claims. Our results demonstrate that Mexico's access-to-information system—originally designed to serve expert and non-expert claims alike—has evolved into a "brokered" system in which experts' claims are more frequent and receive better responses. This transformation owes both to compositional change, as expert claimants have come to make up an increasing portion of total requests, and to behavioral change, as both claimants and obligants have interactively adapted their modes of requesting and providing information.

While the increasing prioritization of expert claimants would appear to exacerbate class bias in Mexican accountability politics, there is perhaps a more optimistic conclusion to draw from this tendency. A request-based ATI system was perhaps never an ideal institutional design to foment widespread participation by ordinary citizens. Rather, one of the most significant accomplishments of Mexico's ATI system has been the development of a large civil society ecosystem that has developed sophisticated tools to use the ATI system productively. Many of these organizations dedicate themselves to supporting grassroots groups throughout the country in accessing information necessary to engage in advocacy around local issues. Strengthening these information brokerage networks would likely be a strong step in favor of accountability. On the other hand, the more routine informational needs, such as finding requests for bids for government contracts or performance indicators of local schools could be better met through proactive information systems, under a "targeted transparency" or "open government" model,[122] rather than a request-based system.

These forms of gradual change should be visible in other participatory institutions of social accountability around the world as well. As anti-corruption agencies or ombudspersons demonstrate better responsiveness to certain types of allegations or complaints, the body of potential claimants will adapt in behavior and potentially evolve in composition as well. As institutions like petitions, social audits, policy councils, participatory budgeting, and local assemblies demonstrate more or less utility for different types of claims or policy areas, different types of participants will be more or less likely to find them worthwhile and may learn new strategies over time as well. Government obligants, in turn, also learn and adapt in response to these changes. We suggest that, over time, these endogenous forms of gradual change can result in institutions that are fundamentally different in their social and political functions, even independent of formal rule changes or politically motivated sabotage.

However, future research should also explore the scope conditions that shape the generalizability of our findings to other countries and other types of social-accountability institutions. We suspect that ATI systems are particularly prone to the gradual and informal processes of change that we

---

[122] Fung 2013

identify because of the large scale of citizen-government interaction, providing repeated engagement through which both claimants and obligants can learn and adjust their behavior. Perhaps more time-consuming—and thus more rarely used—institutions such as anti-corruption bodies may be subject more to external forces rather than internal forces of change. Another trait of ATI systems is that they occupy an intermediate space in terms of how threatening they are to government agents. Given the possibility of disclosing damaging information in response to an information request, agency personnel certainly have some incentive to be non-compliant, particularly when they suspect that a request comes from a political opponent or investigative journalist.[123] Some modes of social-accountability institutions are much less threatening to government agents—such as policy councils—while others are much more threatening, like anti-corruption commissions. We would expect more threatening forms of institutions to prompt particularly acute processes of obligant learning to avoid compliance. Institutional incentives to promote responsiveness are particularly important in these cases. We would also be eager to see a replication of our approach outside of Mexico. The Mexican context was perhaps particularly conducive to the trend toward brokerage given a highly developed ecosystem of civil society organizations that specialize in political accountability.

More recently in Mexico, INAI itself has come under external threat as outgoing President Andrés Manuel López Obrador called for it to be eliminated due to purported ineffectiveness and bloated budget (Villa y Caña and Dina 2023). In light of these attacks, our results suggest that endogenous changes in the social function of institutions may also render them more vulnerable to external political threats. Our results, however, also highlight the sheer scale of information provision that INAI and its predecessor IFAI have managed over two decades, serving a broad and diverse array of users and purposes across society. Were this institution to be eliminated without a sufficiently effective and independent alternative, an immense social demand for information of all kinds would be left unfulfilled.

More broadly, our findings hold two important lessons for future scholarship. First, scholars of participatory and accountability institutions in transitional democracies should move past a narrow concern with adoption and immediate performance, taking seriously the gradual forces that alter the course of these institutions over time. Institutions evolve, both because of environmental shifts and endogenous processes. Such change is felicitous when it yields stronger and more useful tools for social accountability, in a process comparable to institutional lock-in for welfare programs.[124] However, evolution that renders these institutions defunct or introduces bias may be more common. Policy design could potentially take such evolutionary processes into account from the start. For instance, in contexts where professionalization is a likely outcome—as we have identified in the present case—reformers may empower a community of civil society intermediaries to aid less-expert actors in navigating the increasingly sophisticated institution.

---

[123] Gill and Hughes 2005; Berliner et al. 2021; Erlich, Berliner, et al. 2021
[124] Hacker 2004

Data

Replication files for this article can be found at https://doi.org/10.7910/DVN/0LRIAP.

References

Ahmad, Junaid, Stephen Commins, Shantayanan Devarajan, Deon Filmer, Jeffrey Hammer, Lant Pritchett, Ritva Reinikka, Shekhar Shah, and Agnès Soucat. 2003. "World Development Report 2004: Making Services Work for Poor People." The World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/527371468166770790/world-development-report-2004-making-services-work-for-poor-people-overview

Almanzar, Tanya, Mark Aspinwall, and David Crow. 2018. "Freedom of Information in Times of Crisis: The Case of Mexico's War on Drugs." *Governance* 31 (2): 321–39. https://doi-org.libproxy2.usc.edu/10.1111/gove.12289

Arkedis, Jean, Jessica Creighton, Akshay Dixit, Archon Fung, Stephen Kosack, Dan Levy, and Courtney Tolmie. 2021. "Can Transparency and Accountability Programs Improve Health? Experimental Evidence from Indonesia and Tanzania." *World Development* 142. https://doi.org/10.1016/j.worlddev.2020.105369

Baiocchi, Gianpaolo, Patrick Heller, Marcelo Kunrath Silva, and Marcelo Silva. 2011. *Bootstrapping Democracy: Transforming Local Governance and Civil Society in Brazil*. Stanford University Press. https://doi.org/10.1515/9780804777797

Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy* 2 (1): 1–30. https://doi.org/10.1257/pol.2.1.1

Baniamin, Hasan Muhammad, and Ishtiaq Jamil. 2018. "Dynamics of Corruption and Citizens' Trust in Anti-Corruption Agencies in Three South Asian Countries." *Public Organization Review* 18 (3): 381–98. https://doi.org/10.1007/s11115-017-0384-4

Baumgartner, Frank R, and Beth L Leech. 1998. *Basic Interests: The Importance of Groups in Politics and in Political Science*. Princeton University Press. https://doi-org/10.2307/j.ctt7rxwh

Bächtiger, André, and Dominik Hangartner. 2010. "When Deliberative Theory Meets Empirical Political Science: Theoretical and Methodological Challenges in Political Deliberation." *Political Studies* 58: 609–29. https://doi.org/10.1111/j.1467-9248.2010.00835.x

Berliner, Daniel, Benjamin E. Bagozzi, and Brian Palmer-Rubin. 2018. “What Information Do Citizens Want? Evidence from One Million Information Requests in Mexico.” *World Development* 109: 222–35. https://doi.org/10.1016/j.worlddev.2018.04.016

Berliner, Daniel, Benjamin E. Bagozzi, Brian Palmer-Rubin, and Aaron Erlich. 2021. “The Political Logic of Government Disclosure: Evidence from Information Requests in Mexico.” *Journal of Politics* 81 (1): 229–45. https://doi.org/10.1086/709148

Berliner, Daniel, and Aaron Erlich. 2015. “Competing for Transparency: Political Competition and Institutional Reform in Mexican States.” *American Political Science Review* 109 (1): 110–28. https://doi.org/10.1017/S0003055414000616

Berliner, Daniel, Brian Palmer-Rubin, Jésica E. Tapia Reyes, Benjamin E. Bagozzi, and Aaron Erlich. 2022. “Big Data y Acceso a la Información en México.” London School of Economics and Political Science. https://osf.io/preprints/osf/nbqus

Besley, Timothy, and Robin Burgess. 2002. “The Political Economy of Government Responsiveness: Theory and Evidence from India.” *The Quarterly Journal of Economics* 117 (4): 1415–51. https://doi.org/10.1162/003355302320935061

Björkman, Martina, and Jakob Svensson. 2009. “Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda.” *The Quarterly Journal of Economics* 124 (2): 735–69. https://doi.org/10.1162/qjec.2009.124.2.735

Bookman, Zachary, and Juan-Pablo Guerrero Amparán. 2009. “Two Steps Forward, One Step Back: Assessing the Implementation of Mexico’s Freedom of Information Act.” *Mexican Law Review* 1 (2): 25–49. https://revistas.juridicas.unam.mx/index.php/mexican-law-review/article/view/7721/9689

Buntaine, Mark T, Patrick Hunnicutt, and Polycarp Komakech. 2021. “The Challenges of Using Citizen Reporting to Improve Public Services: A Field Experiment on Solid Waste Services in Uganda.” *Journal of Public Administration Research and Theory* 31 (1): 108–27. https://doi.org/10.1093/jopart/muaa026

Buntaine, Mark T, Daniel L Nielson, and Jacob T Skaggs. 2021. “Escaping the Disengagement Dilemma: Two Field Experiments on Motivating Citizens to Report on Public Services.” *British Journal of Political Science* 51 (2): 685–705.

Call, Charles, and Jeffrey Hallock. 2020. “Too Much Success? The Legacy and Lessons of the International Commission Against Impunity in Guatemala.” American University CLALS Working Paper Series No. 24. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3526865

Cameron, Maxwell A. 2020. “The Return of Oligarchy? Threats to Representative Democracy in Latin America.” *Third World Quarterly* 42 (4): 775–92. https://doi.org/10.1080/01436597.2020.1865794

Carpenter, Daniel. 2021. *Democracy by Petition: Popular Politics in Transformation, 1790-1870*. Cambridge, Mass.: Harvard University Press. https://doi.org/10.4159/9780674258921

Cejudo, Guillermo M, Sergio López Ayllón, and Alejandra Ríos Cázares. 2014. *La Política de Transparencia en México: Instituciones, Logros y Desafíos*. Mexico City: Editorial CIDE.

Cejudo, Guillermo M, Alejandra Ríos, and Sergio López-Ayllón. 2010. "Métrica de La Transparencia 2010." *Centro de Investigación y Docencia Económicas*. http://metricadetransparencia2010.cide.edu/Metrica%202010.pdf

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A Scalable Tree Boosting System." In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–94. https://doi.org/10.1145/2939672.2939785

Corbacho, Ana, Daniel W. Gingerich, Virginia Oliveros, and Mauricio Ruiz-Vega. 2016. "Corruption as a Self-Fulfilling Prophecy: Evidence from a Survey Experiment in Costa Rica." *American Journal of Political Science* 60 (4): 1077–92. https://doi.org/10.1111/ajps.12244

Crepaz, Michele. 2020. "To Inform, Strategise, Collaborate, or Compete: What Use Do Lobbyists Make of Lobby Registers?" *European Political Science Review* 12 (3): 347–69. https://doi.org/10.1017/S1755773920000156

Dipoppa, Gemma, and Guy Grossman. 2020. "The Effect of Election Proximity on Government Responsiveness and Citizens' Participation: Evidence from English Local Elections." *Comparative Political Studies* 53 (14): 2183–2212. https://doi.org/10.1177/0010414020912290

Distelhorst, Greg, and Yue Hou. 2017. "Constituency Service Under Nondemocratic Rule: Evidence from China." *The Journal of Politics* 79 (3): 1024–40. https://doi.org/10.1086/690948

Donaghy, Maureen M. 2020. "Participation Meets Politics: Political Shifts and the Longevity of Participatory Governance Institutions." *PS: Political Science & Politics* 53 (1): 6–9. https://doi.org/10.1017/S1049096519001331

Dunning, Thad, and others. 2019. "Voter Information Campaigns and Political Accountability: Cumulative Findings from a Preregistered Meta-Analysis of Coordinated Trials." *Science Advances* 5 (7). https://doi.org/10.1126/sciadv.aaw2612

Ebrahim, Alnoor, Julie Battilana, and Johanna Mair. 2014. "The Governance of Social Enterprises: Mission Drift and Accountability Challenges in Hybrid Organizations." *Research in Organizational Behavior* 34: 81–100. https://doi.org/10.1016/j.riob.2014.09.001

Einstein, Katherine Levine, Maxwell Palmer, and David M Glick. 2019. “Who Participates in Local Government? Evidence from Meeting Minutes.” *Perspectives on Politics* 17 (1): 28–46. https://doi-org/10.1017/S153759271800213X

Erlich, Aaron, Daniel Berliner, Brian Palmer-Rubin, and Benjamin E Bagozzi. 2021. “Media Attention and Bureaucratic Responsiveness.” *Journal of Public Administration Research and Theory* 31 (4): 687–703. https://doi-org/10.1093/jopart/muab001

Erlich, Aaron, Stefano G Dantas, Benjamin E Bagozzi, Daniel Berliner, and Brian Palmer-Rubin. 2021. “Multi-Label Prediction for Political Text-as-Data.” *Political Analysis* 30 (4): 463–80. https://doi.org/10.1017/pan.2021.15

Falleti, Tulia G, and Thea N Riofrancos. 2018. “Endogenous Participation: Strengthening Prior Consultation in Extractive Economies.” *World Politics* 70 (1): 86–121. https://doi.org/10.1017/S004388711700020X

Ferraz, Claudio, and Frederico Finan. 2008. “Exposing Corrupt Politicians: The Effects of Brazil’s Publicly Released Audits on Electoral Outcomes.” *The Quarterly Journal of Economics* 123 (2): 703–45. https://doi.org/10.1162/qjec.2008.123.2.703

Fong, Christian, and Matthew Tyler. 2021. “Machine Learning Predictions as Regression Covariates.” *Political Analysis* 29 (4): 467–84. https://doi.org/10.1017/pan.2020.38.

Fox, Jonathan. 2007. *Accountability Politics: Power and Voice in Rural Mexico*. New York: Oxford University Press.

Fox, Jonathan. 2015. “Social Accountability: What Does the Evidence Really Say?” *World Development* 72: 346–61. https://doi.org/10.1016/j.worlddev.2015.03.011

Fox, Jonathan, Libby Haight, and Brian Palmer-Rubin. 2011. “Delivering Transparency: To What Degree Does the Mexican Government Respond to Public Information Requests.” *Gestión y Política Pública* 20 (1): 3–61.

Fung, Archon. 2013. “Infotopia: Unleashing the Democratic Power of Transparency.” *Politics & Society* 41 (2): 183–212. https://doi.org/10.1177/0032329213483107

Gaventa, John, and Gregory Barrett. 2012. “Mapping the Outcomes of Citizen Engagement.” *World Development* 40 (12): 2399–2410. https://doi.org/10.1016/j.worlddev.2012.05.014

Gill, Juliet, and Sallie Hughes. 2005. “Bureaucratic Compliance with Mexico’s New Access to Information Law.” *Critical Studies in Media Communication* 22 (2): 121–37. https://doi.org/10.1080/07393180500072038

Goetz, Anne Marie, and Rob Jenkins. 2001. “Hybrid Forms of Accountability: Citizen Engagement in Institutions of Public-Sector Oversight in India.” *Public Management Review* 3 (3): 363–83. https://doi.org/10.1080/14616670110051957

Grossman, Guy, Melina R Platas, and Jonathan Rodden. 2018. “Crowdsourcing Accountability: ICT for Service Delivery.” *World Development* 112: 74–87. https://doi.org/10.1016/j.worlddev.2018.07.001

Grzymala-Busse, Anna. 2006. “The Discrete Charm of Formal Institutions: Postcommunist Party Competition and State Oversight.” *Comparative Political Studies* 39 (3): 271–300. https://doi.org/10.1177/0010414005284216

Hacker, Jacob S. 2004. “Privatizing Risk Without Privatizing the Welfare State: The Hidden Politics of Social Policy Retrenchment in the United States.” *American Political Science Review* 98 (2): 243–60. https://doi.org/10.1017/S0003055404001121

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. “How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice.” *Political Analysis* 27 (2): 163–92. https://doi.org/10.1017/pan.2018.46

Hossain, Naomi, Anuradha Joshi, and Suchi Pande. 2024. “The Politics of Complaint: A Review of the Literature on Grievance Redress Mechanisms in the Global South.” *Policy Studies* 45 (2), 139–58. https://doi.org/10.1080/01442872.2023.2193387

INEGI. 2020. “Encuesta Nacional de Acceso a la Información Pública y Protección de Datos Personales 2019 (ENAID): Principales Resultados.” https://www.inegi.org.mx/programas/enaid/2019/

Joshi, Anuradha, and Peter P. Houtzager. 2012. “Widgets or Watchdogs? Conceptual Explorations in Social Accountability.” *Public Management Review* 14 (2): 145–62. https://doi.org/10.1080/14719037.2012.657837

Kim, Jieun, Rachel E Stern, Benjamin L Liebman, and Xiaohan Wu. 2021. “Closing Open Government: Grassroots Policy Conversion of China’s Open Government Information Regulation and Its Aftermath.” *Comparative Political Studies* 55 (2), 319–47. https://doi.org/10.1177/00104140211024314

Kosack, Stephen, Courtney Tolmie, and Charles C Griffin. 2010. *From the Ground Up: Improving Government Performance with Independent Monitoring Organizations*. Brookings Institution Press.

Kruks-Wisner, Gabrielle. 2021. “Great Expectations, Great Grievances: The Politics of Citizens’ Complaints in India.” *Comparative Politics* 54 (1): 27–49. https://doi.org/10.5129/001041521X16106633880914

Kruks-Wisner, Gabrielle. 2022. “Social Brokerage: Accountability and the Social Life of Information.” *Comparative Political Studies* 55 (14): 1–34. https://doi.org/10.1177/00104140221089644

Kwoka, Margaret B. 2015. “FOIA, Inc.” *Duke Law Journal* 65: 1361.

Lachenel, Cécile, and Ana Joaquina Ruiz Guerra. 2013. *Derecho de Acceso a La Información. La Visión de Los Usuarios*. Mexico City: Fundar, Centro de Análisis e Investigación, A.C.; Editorial Gedisa.

Lagunes, Paul, and Oscar Pocasangre. 2019. “Dynamic Transparency: An Audit of Mexico’s Freedom of Information Act.” *Public Administration* 97 (1): 162–76. https://doi.org/10.1111/padm.12553

Levitsky, Steven, and María Victoria Murillo. 2009. “Variation in Institutional Strength.” *Annual Review of Political Science* 12: 115–33. https://doi.org/10.1146/annurev.polisci.11.091106.121756

Lieberman, Evan S., Daniel N. Posner, and Lily L. Tsai. 2014. “Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya.” *World Development* 60: 69–83. https://doi.org/10.1016/j.worlddev.2014.03.014

Lindblom, Charles E. 1977. *Politics and Markets: The World’s Political Economic Systems*. New York: Basic Books.

López Aceves, Juan Carlos. 2013. *Transparencia en México: Documentos, Actores y Fechas Claves, 2001-2013*. Mexico City, Mexico: Instituto de Transparencia y Acceso a la Información Pública de Tamaulipas.

Lührmann, Anna, Kyle L Marquardt, and Valeriya Mechkova. 2020. “Constraining Governments: New Indices of Vertical, Horizontal, and Diagonal Accountability.” *American Political Science Review* 114 (3): 811–20. https://doi.org/10.1017/S0003055420000222

Mahoney, James, and Kathleen Thelen. 2010. “A Theory of Gradual Institutional Change.” In James Mahoney and Kathleen Thelen, eds., *Explaining Institutional Change: Ambiguity, Agency, and Power*. New York: Cambridge University Press, pp. 1-37.

Forster, Reinder, Carmen Malena, and Janmejay Singh. 2004. “Social Accountability: An Introduction to the Concept and Emerging Practice.” Social Development Papers No. 76. Washington, D.C.: World Bank. http://documents.worldbank.org/curated/en/327691468779445304/Social-accountability-an-introduction-to-the-concept-and-emerging-practice

Mansuri, Ghazala, and Vijayendra Rao. 2013. “Localizing Development: Does Participation Work?” Washington, D.C.: World Bank. https://www.worldbank.org/en/research/publication/localizing-development-does-participation-work

Mayka, Lindsay. 2019. *Building Participatory Institutions in Latin America: Reform Coalitions and Institutional Change*. New York: Cambridge University Press. https://doi.org/10.1017/9781108598927

Mayka, Lindsay, and Jared Abbott. 2023. "Varieties of Participatory Institutions and Interest Intermediation." *World Development* 171: 106369. https://doi.org/10.1016/j.worlddev.2023.106369

McGee, Rosie, and John Gaventa. 2011. "Shifting Power? Assessing the Impact of Transparency and Accountability Initiatives." IDS Working Paper No. 383. https://doi.org/10.1111/j.2040-0209.2011.00383_2.x

Merino, Mauricio. 2013. "El IFAI: En Busca del Ejercicio Cotidiano de la Democracia." In *10 Años de Transparencia en México*. Instituto Federal de Acceso a la Información y Protección de Datos México, pp. 73–94.

Michener, Gregory. 2011. "FOI Laws Around the World." *Journal of Democracy* 22 (2): 145–59. https://doi.org/10.1353/jod.2011.0021

Michener, Gregory. 2013. "La Libertad de Información. Una Institución 'Boutique' en México." In Cécile Lachenel and Ana Joaquina Ruiz Guerra, eds., *Derecho de Acceso a La Información. La Visión de Los Usuarios*. Mexico City: Fundar, Centro de Análisis e Investigación, A.C.; Editorial Gedisa, pp. 83–101.

Michener, Gregory, Evelyn Contreras, and Irene Niskier. 2018. "From Opacity to Transparency? Evaluating Access to Information in Brazil Five Years Later." *Revista de Administração Pública* 52 (4): 610–29.

Moynihan, Donald, Pamela Herd, and Hope Harvey. 2015. "Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions." *Journal of Public Administration Research and Theory* 25 (1): 43–69. https://doi.org/10.1093/jopart/muu009

Mungiu-Pippidi, Alina. 2015. *The Quest for Good Governance: How Societies Develop Control of Corruption*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9781316286937

O'Donnell, Guillermo. 1998. "Horizontal Accountability in New Democracies." *Journal of Democracy* 9 (3): 112–26. https://doi.org/10.1353/jod.1998.0051

Olaizola, Beatriz, and Daniele Grasso. 2023. "Decide Madrid, la plataforma de participación ciudadana que el gobierno de Almeida dejó morir." *El Pais*. July 26, 2023. https://elpais.com/espana/madrid/2023-07-26/decide-madrid-la-plataforma-de-participacion-ciudadana-que-el-gobierno-de-almeida-dejo-morir.html

Palmer-Rubin, Brian. 2022. *Evading the Patronage Trap: Interest Representation in Mexico*. University of Michigan Press. https://doi.org/10.3998/mpub.12132850

Palmer-Rubin, Brian; Tapia Reyes, Jésica E.; Berliner, Daniel; Erlich, Aaron; Bagozzi, Benjamin E., 2024, "Replication Data for: Accountability in Time: Evolution in Social-Accountability Institutions." Harvard Dataverse. https://doi.org/10.7910/DVN/0LRIAP

Persson, Anna, Bo Rothstein, and Jan Teorell. 2013. “Why Anticorruption Reforms Fail: Systemic Corruption as a Collective Action Problem.” *Governance* 26 (3): 449–71. https://doi.org/10.1111/j.1468-0491.2012.01604.x

Pierson, Paul. 1993. “When Effect Becomes Cause: Policy Feedback and Political Change.” *World Politics* 45 (4): 595–628. https://doi.org/10.2307/2950710

Pierson, Paul. 2011. *Politics in Time: History, Institutions, and Social Analysis*. Princeton University Press. https://doi.org/10.1515/9781400841080

Piñeiro Rodríguez, Rafael, Paula Muñoz, Fernando Rosenblatt, Cecilia Rossel, and Fabrizio Scrollini. 2021. “Domestic Isomorphic Pressures in the Design of FOI Oversight Institutions in Latin America.” *Governance* 35 (3): 827–47. https://doi.org/10.1111/gove.12614

PlanDAI. 2020. “Plan Nacional de Socialización del Derecho de Acceso a la Información.” https://micrositios.inai.org.mx/plandai/

Relly, Jeannine E, Md Fazle Rabbi, Meghna Sabharwal, Rajdeep Pakanati, and Ethan H Schwalbe. 2020. “More Than a Decade in the Making: A Study of the Implementation of India’s Right to Information Act.” *World Development* 136: 105088. https://doi.org/10.1016/j.worlddev.2020.105088

Ríos Cázares, Alejandra, Carmen Castañeda, and Adriana García. 2017. “Transparencia. Diagnóstico Institucional 2016.” Mexico City: CIDE-INAI.

Robinson, Mark. 2006. *Budget Analysis and Policy Advocacy: The Role of Non-Governmental Public Action*. IDS Working Paper No. 279.

Rothstein, Bo. 2011. “Anti-Corruption: The Indirect ‘Big Bang’ Approach.” *Review of International Political Economy* 18 (2): 228–50. https://doi.org/10.1080/09692291003607834

Sanyal, Paromita, and Vijayendra Rao. 2019. *Oral Democracy: Deliberation in Indian Village Assemblies*. New York: Cambridge University Press.

Schedler, Andreas. 1999. “Conceptualizing Accountability.” In Andreas Schedler, Larry Jay Diamond, and Marc F Plattner, eds., *The Self-Restraining State: Power and Accountability in New Democracies*. Boulder, Col.: Lynne Rienner Publishers, pp. 13–28.

Shim, Dong Chul, and Tae Ho Eom. 2008. "E-Government and Anti-Corruption: Empirical Analysis of International Data." *International Journal of Public Administration* 31 (3): 298–316. https://doi.org/10.1080/01900690701590553

Shwartz-Ziv, Ravid, and Amitai Armon. 2021. "Tabular Data: Deep Learning Is Not All You Need." In *8th ICML Workshop on Automated Machine Learning (AutoML)*.

Sjoberg, Fredrik M, Jonathan Mellon, and Tiago Peixoto. 2017. "The Effect of Bureaucratic Responsiveness on Citizen Participation." *Public Administration Review* 77 (3): 340–51. https://doi.org/10.1111/puar.12697

Slough, Tara, Daniel Rubenson, Ro'ee Levy, Francisco Alpizar Rodriguez, María Bernedo del Carpio, Mark T Buntaine, Darin Christensen, et al. 2021. "Adoption of Community Monitoring Improves Common Pool Resource Management Across Contexts." *Proceedings of the National Academy of Sciences* 118 (29): e2015367118. https://doi.org/10.1073/pnas.2015367118

Smulovitz, Catalina, and Enrique Peruzzotti. 2000. "Societal Accountability in Latin America." *Journal of Democracy* 11 (4): 147–58. https://doi.org/10.1353/jod.2000.0087

Sobel, David, Bethany Davis Noll, Benjamin Fernandez Bogado, TCG Group, and Monroe Price. 2006. "The Federal Institute for Access to Information and a Culture of Transparency." Annenberg School of Communications, Philadelphia: University of Pennsylvania.

Steiner, Jürg, André Bächtiger, Markus Spörndli, and Marco R. Steenbergen. 2004. *Deliberative Politics in Action: Analysing Parliamentary Discourse*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511491153

Steinsson, Sverrir. 2024. "Rule Ambiguity, Institutional Clashes, and Population Loss: How Wikipedia Became the Last Good Place on the Internet." *American Political Science Review* 118 (1): 235-51. https://doi.org/10.1017/S0003055423000138

Tang, Wan, Jun Hu, Hui Zhang, Pan Wu, and Hua He. 2015. "Kappa Coefficient: A Popular Measure of Rater Agreement." *Shanghai Archives of Psychiatry* 27 (1): 62-67. https://doi.org/10.11919%2Fj.issn.1002-0829.215010

Tarrow, Sidney. 2005. *The New Transnational Activism*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511791055

Thelen, Kathleen. 2003. "How Institutions Evolve: Insights from Historical Institutional Analysis." In James Mahoney and Dietrich Rueschemeyer, eds., *Comparative Historical Analysis in the Social Sciences*. New York: Cambridge University Press, pp. 208–240.

UNESCO. 2022. "Global Report on the Implementation of Access to Information Laws." https://unesdoc.unesco.org/ark:/48223/pf0000383160.locale=en

Verba, Sidney, Kay Lehman Schlozman, and Henry E Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, Mass.: Harvard University Press.

Wampler, Brian. 2015. *Activating Democracy in Brazil: Popular Participation, Social Justice, and Interlocking Institutions*. Notre Dame, IN: University of Notre Dame Press.

Wampler, Brian, Natasha Borges Sugiyama, and Michael Touchton. 2019. *Democracy at Work: Pathways to Well-Being in Brazil*. New York: Cambridge University Press. https://doi.org/10.1017/9781108675949

Wille, Anchrit, and Mark Bovens. 2022. "Watching EU Watchdogs Assessing the Accountability Powers of the European Court of Auditors and the European Ombudsman." *Journal of European Integration* 44 (3): 183–206. https://doi.org/10.1080/07036337.2020.1852229

Worthy, Ben. 2010. "More Open but Not More Trusted? The Effect of the Freedom of Information Act 2000 on the United Kingdom Central Government." *Governance* 23 (4): 561–82. https://doi.org/10.1111/j.1468-0491.2010.01498.x

Yang, Kaifeng, and Kathe Callahan. 2007. "Citizen Involvement Efforts and Bureaucratic Responsiveness: Participatory Values, Stakeholder Pressures, and Administrative Practicality." *Public Administration Review* 67 (2): 249–64. https://doi.org/10.1111/j.1540-6210.2007.00711.x

Acknowledgements

We thank Carmen Castañeda Farill, Stefano Dantas, Manuel Gutiérrez, Sandra Juan Delgado, Katia Guzmán Martínez, Edith Mercado Gómez, Andrea Sánchez Mercado, and Oleg Zhilin for their research assistance and Diego Díaz Iturbe, Lourdes Antonioli Ortíz, and Jaime Cerdio Moisés at INAI for help accessing data. Palmer-Rubin's research was supported by the Marquette University Committee on Research. Berliner's research was supported by the Arizona State University School of Politics and Global Studies and Program for Transborder Communities; and by the London School of Economics Department of Government, LSE Knowledge Exchange and Impact Fund, and the LSE Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD). Erlich's research was made possible through Social Science and Humanities Research Council (SSHRC) Grant #430-2018-1069, Fonds de recherche du Québec – Société et culture (FQRSC) Grant #253243, and Compute Canada. Bagozzi's research was partially supported by the National Science Foundation under Grant No. DMS-1737865. For feedback on earlier versions of this project, the authors thank Taylor Boas, Ana De la O, Jonathan Fox, ShahBano Igaz, Ryan Jablonski, Milli Lake, Tiffany Lau, Miguel Rueda, Brigitte Seim and participants in panels or presentations at the following meetings: 2021 Latin American Studies Association, 2021 Midwest Political Science Association, 2021 American Political Science Association, 2022 Global Conference on Transparency Research, 2022 London Public Policy Mini-Conference, and 2024 SoCal Political Institutions and Political Economy Conference, as well as three anonymous reviewers.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Year	0.0036^{***}	0.0037^{***}	0.0033^{***}	0.0030^{***}	0.0028^{***}	0.0029^{***}	0.0029^{***}
	(0.0006)	(0.0005)	(0.0003)	(0.0004)	(0.0005)	(0.0003)	(0.0003)
Num. obs.	1,166,727	1,166,727	1,166,727	1,166,727	1,166,727	1,166,727	1,166,727
Adj. R^2	0.0117	0.1062	0.2586	0.2356	0.3934	0.3690	0.4910
FE: Municipality		2,074					
FE: Postcode			27,275				
FE: Muni × Agency				60,188			
FE: Postcode × Agency					287,811		
FE: Muni × Agency × Theme						122,309	
FE: Postcode × Agency × Theme							436,251

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 1: Linear models of request expertise, for requests 2003-2016. Standard errors clustered by municipality. For each set of fixed effects, numbers indicate the number of distinct groups.

	Model 1	Model 2	Model 3	Model 4	Model 5
Dependent Variable	Log Days-to-Response	Official Inexistencia	True Inexistencia	Uninformative	Noncompliant
Expert Request Score	1.181^{***}	0.127^{***}	0.126^{***}	0.155	0.134^{***}
	(0.165)	(0.036)	(0.042)	(0.098)	(0.048)
Years	0.024^{***}	-0.000	0.010^{***}	-0.005^{**}	0.006^{***}
	(0.004)	(0.001)	(0.001)	(0.002)	(0.002)
Expert Request Score × Years	-0.115^{***}	-0.012^{***}	-0.019^{***}	-0.007	-0.007^{*}
	(0.019)	(0.003)	(0.003)	(0.008)	(0.004)
Adj. R^2	0.174	0.085	0.062	0.108	0.068
Num. obs.	$1,610,997$	$1,611,289$	$1,537,385$	$1,537,385$	$924,920$
Controls	Y	Y	Y	Y	Y
Agency Fixed Effects	Y	Y	Y	Y	Y

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

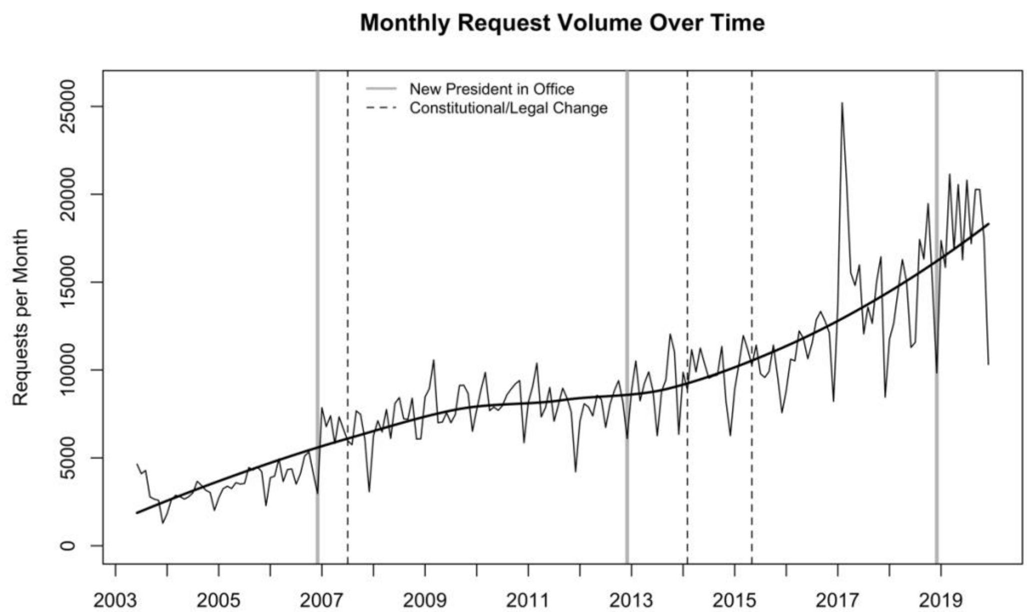
Table 2: Linear models of request-level poor responsiveness. All models include agency fixed effects and controls for the month-of-year, each request's logged word count, readability, medium, inclusion of an attachment, as well as predicted measures of request appropriateness, complexity, and theme. Standard errors clustered by agency.

Dependent Variable	Model 1 Log Days-to-Response	Model 2 Official Inexistencia	Model 3 True Inexistencia	Model 4 Uninformative	Model 5 Noncompliant
Expert Request Score	1.374^{***}	0.113^{*}	0.306^{***}	0.071	0.291^{***}
	(0.378)	(0.059)	(0.074)	(0.111)	(0.079)
Agency Lagged Avg. Expertise	1.052^{**}	-0.258^{***}	0.840^{***}	-0.742^{***}	0.527^{*}
	(0.494)	(0.086)	(0.179)	(0.208)	(0.283)
Expert Request Score × Agency Lagged Avg. Expertise	-7.199^{***}	-0.593^{***}	-1.880^{***}	0.049	-1.142^{***}
	(1.851)	(0.221)	(0.351)	(0.573)	(0.408)
Adj. R^2	0.173	0.085	0.059	0.107	0.066
Num. obs.	$1{,}577{,}378$	$1{,}577{,}605$	$1{,}507{,}203$	$1{,}507{,}203$	$909{,}650$
Controls	Y	Y	Y	Y	Y
Agency Fixed Effects	Y	Y	Y	Y	Y

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 3: Linear models of request-level poor responsiveness. All models include agency fixed effects and controls for the month-of-year, each request's logged word count, readability, medium, inclusion of an attachment, as well as predicted measures of request appropriateness, complexity, and theme. Standard errors clustered by agency.

Figures



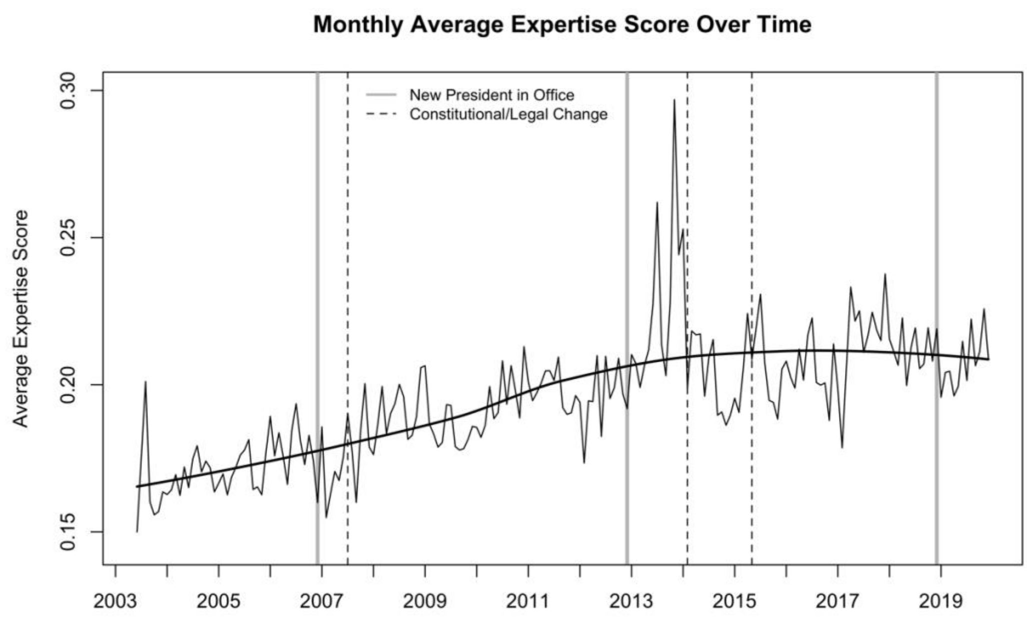


Figure 1: Trends over time in overall volume of information requests (panel 1) and in the average expertise score of information requests (panel 2), both by month. In each panel, the bold line is a smoothed trend over time. Vertical grey lines mark new presidents taking office. Vertical dashed lines mark the adoption of new constitutional or legal changes to Mexico's ATI regime: Reforms to Article 6 of Mexico's constitution in July 2007 and February 2014, and the passage of the General Transparency Law in May 2015.

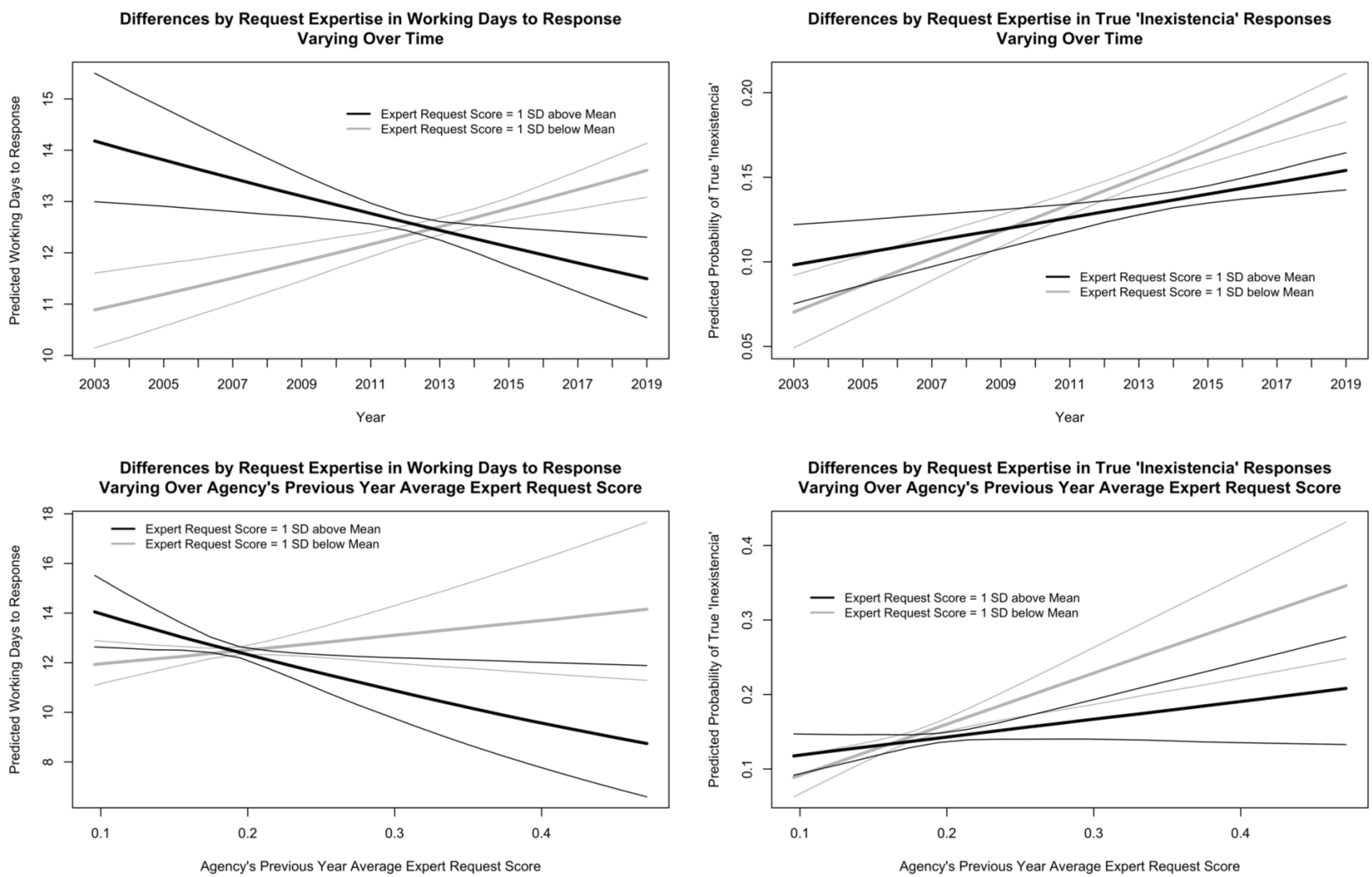


Figure 2: Visualizations of key results from Models 1 and 3 in Table 1 and Models 1 and 3 in Table 2.

Supporting Information: Accountability in Time: Evolution and Expertise in Participatory Institutions

A Variable Definitions

Table A.1: Descriptions of Independent Variables

Variable	Measurement Approach
Expertise Score	Sum of nine request measures; three are language-based and six are specificity-based. The language measures are classifier predictions of whether or not the request used formal, legal, or technical language. The specificity measures are classifier predictions of whether or not the request mentioned specific documents (by name or number), persons (by name or title), dates (more specific than a year), places (more specific than a state), governmental institutions (more specific than the name of the agency; for example a sub-secretariat or internal area), or organizations (private or non-governmental). The sum of the three language measures is divided by three, while the sum of the six language measures is divided by six, before the two sums are added together, so that the overall expertise score is comprised equally of language and specificity measures.
Expertise Score (PCA version)	As a robustness check, we use the first dimension of the results of a principal components analysis of the nine language and specificity measures discussed above.
Year	Measure of time in years (on the date a request is filed) counting 2003 as zero, but varying by month (so that, e.g. February 2009 is 6.083 while March 2009 is 6.167).
Agency Lag Avg. Expert.	Average expertise score of all requests received by a given agency in prior year.
Log Word Count	Logged count of words in the request text, including any attached files.
Readability	Modified readability score of the request text. Since many requests consist of only a single sentence, canonical readability metrics are not applicable here. Our readability score is thus calculated as the ratio of the number of characters to the number of words.
Attachment	An indicator of whether or not the request contained any attached files.
Non-electronic Request	An indicator for any requests filed non-electronically (and thus entered into the system by officials). This only applies to 2.6 percent of all requests.
Inappropriate Request	Sum of four classifier predictions for indicators of requests that may be inappropriate and thus reasonably denied or redirected for legally compliant reasons: Unclear request; requested information not the competency of the agency; requested information likely to be legitimately classified or confidential; and requested information not likely to exist. This variable is particularly important to capture as inappropriate requests (e.g, that can be redirected to another entity) can often be handled very quickly by the agency's Transparency Unit without needing to liaise with other agency administrative units or with the agency's leadership, counsel, or information committee.
Multiple Queries	A measure of request complexity based on classifier predictions of whether or not the request had multiple distinct queries.
Multiple Areas	A measure of request complexity based on classifier predictions of whether the request had queries relating to multiple distinct administrative units of the agency.
Unrelated Queries	A measure of request complexity based on classifier predictions of whether or not the request contained distinct queries that were unrelated to each other.
Info. Type-Datum	Classifier prediction of whether or not the request sought a single data point.
Info. Type-Data	Classifier prediction of whether or not the request sought multiple data points.
Info. Type-Database	Classifier prediction of whether or not the request sought a database.
Info. Type-Document	Classifier prediction of whether or not the request sought a document.
Info. Type-Multiple Docs	Classifier prediction of whether or not the request sought multiple documents.
Info. Theme-Activities	Classifier prediction of whether or not the request pertained to institutional activities.
Info. Theme-Budgets	Classifier prediction of whether or not the request pertained to budgets or spending.
Info. Theme-Evaluation	Classifier prediction of whether or not the request pertained to evaluations or statistics.
Info. Theme-Contracts	Classifier prediction of whether or not the request pertained to external contracts.
Info. Theme-Personnel	Classifier prediction of whether or not the request pertained to institutional structures, personnel, or human resources.
Info. Theme-Regulatory	Classifier prediction of whether the request pertained to regulatory matters or permits.
Info. Theme-Other	Classifier prediction of whether the request's theme did not fall into the above themes.

B Comparison of Request Expertise and Municipality-Level Economic Marginality

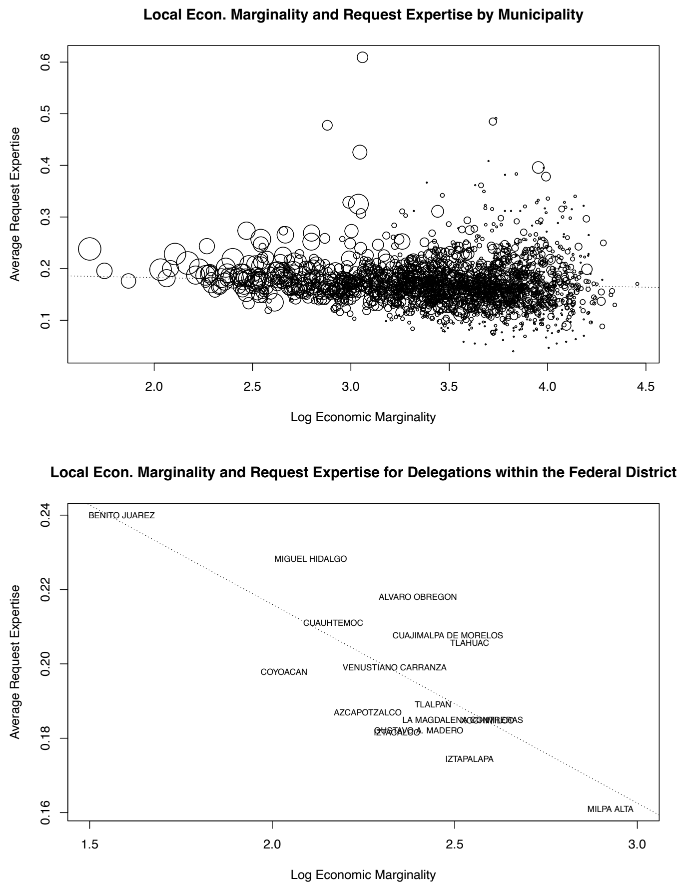


Figure B.1: Comparisons of Request Expertise Score with municipality-level Economic Marginality. The first panel plots the average expertise score of requests from each municipality against the logged 2000 Economic Marginality index of that municipality (with higher values indicating poorer municipalities). Circles are scaled to the logged number of requests filed from each municipality. The bivariate relationship (slope = -0.007) is negative and statistically significant, indicating that more expert requests are filed from wealthier municipalities, and vice versa. This relationship can be seen more clearly, with fewer confounding differences between different types of municipalities and areas of the country, in the second panel, which plots the same relationship but only for the sixteen delegaciones (the equivalent administrative unit to municipalities within the Federal District) of Mexico City. Again the relationship (slope = -0.053) is negative and statistically significant, as more expert requests are filed from wealthier delegaciones.

C Histograms of the relative distributions of expertise score in 2003 and 2019

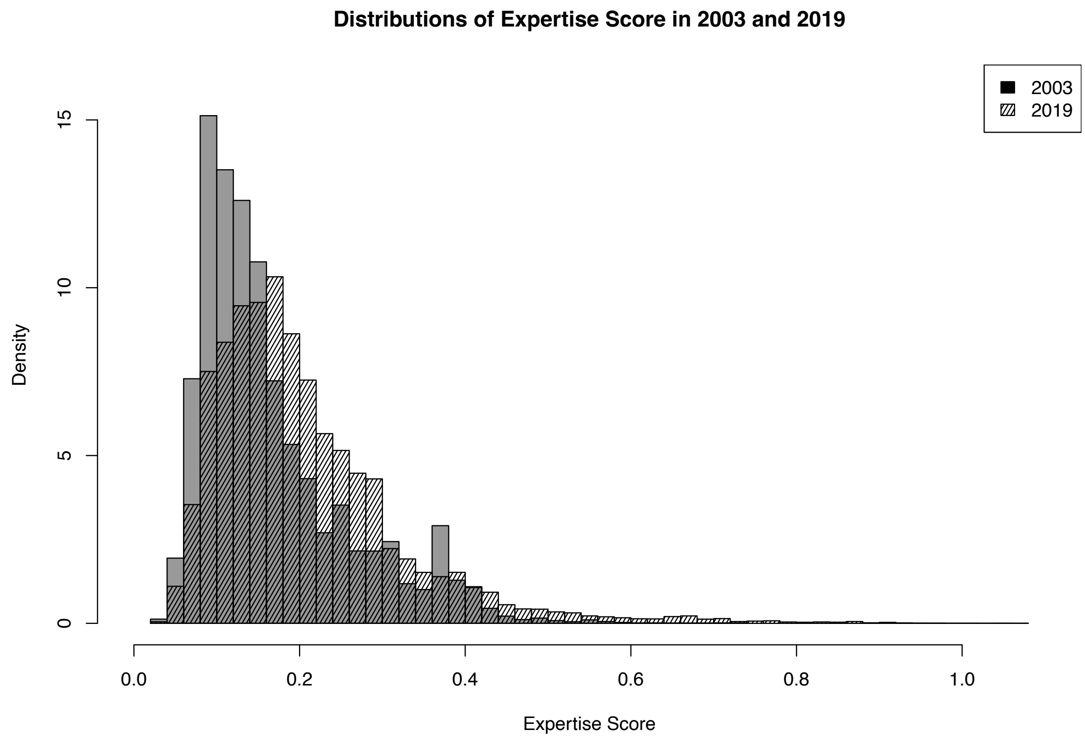


Figure C.2: Histograms of the relative distributions of request expertise score in 2003 (shaded in gray) and 2019 (shaded with diagonal lines).

D Time series plots of measures of responsiveness to information requests

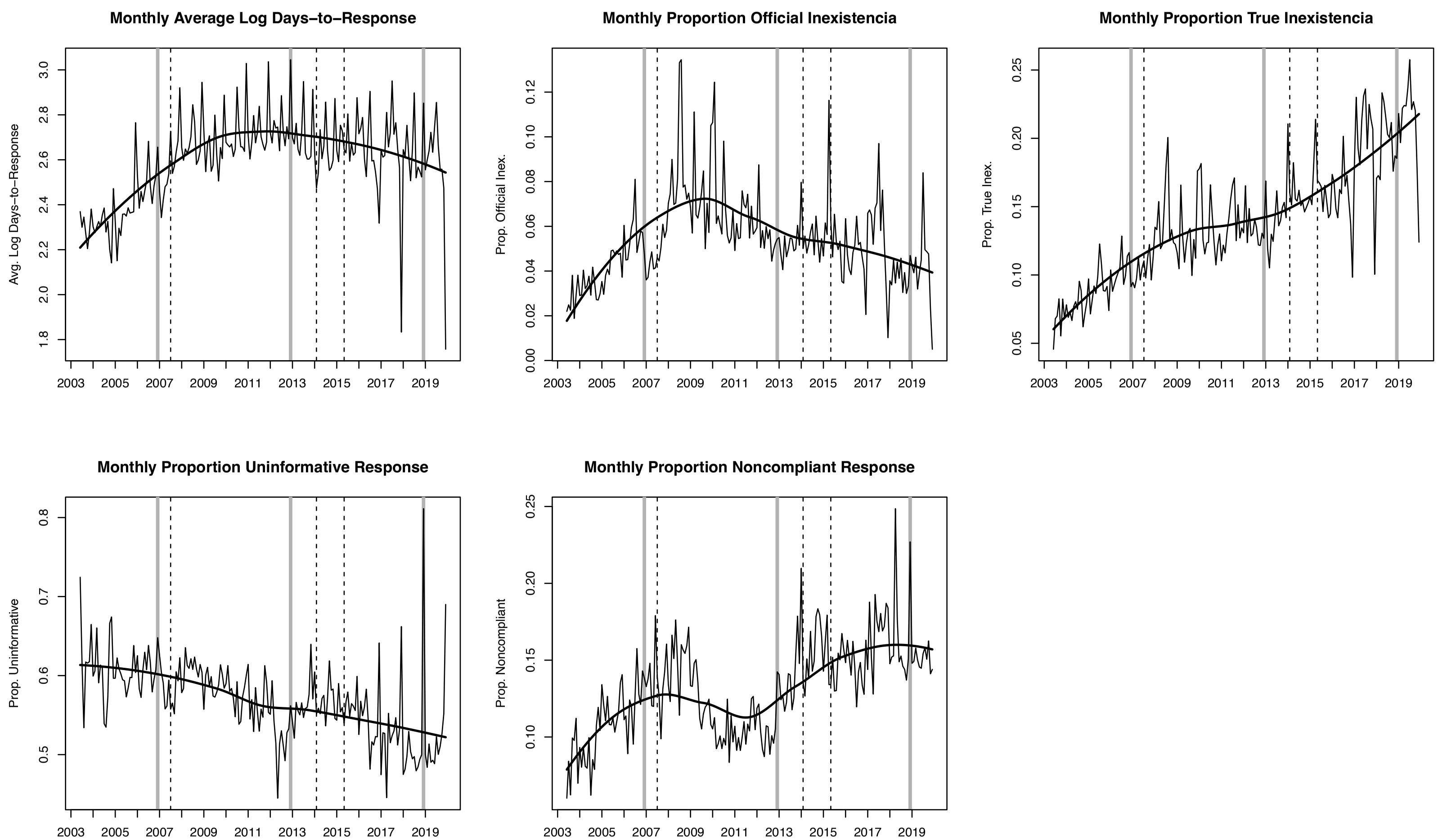


Figure D.3: Trends over time in five different measures of responsiveness to information requests, by monthly averages. In all cases higher values reflect *less* responsiveness. In each panel, the bold line is a smoothed trend over time. Vertical grey lines mark new presidents taking office. Vertical dashed lines mark the adoption of new constitutional or legal changes to Mexico's ATI regime: Reforms to Article 6 of Mexico's constitution in July 2007 and February 2014, and the passage of the General Transparency Law in May 2015.

E Summary statistics

	Min.	Max.	Mean	Median	SD
Log Days-to-Response	0.00	7.58	2.66	2.94	0.91
Official Inexistencia	0.00	1.00	0.06	0.00	0.23
True Inexistencia	0.00	1.00	0.16	0.00	0.36
Uninformative Response	0.00	1.00	0.56	1.00	0.50
Noncompliant Response	0.00	1.00	0.14	0.00	0.35
Expert Request Score	0.02	1.16	0.20	0.17	0.12
Year	0.42	16.92	10.58	11.25	4.42
Agency Lagged Avg. Expertise	0.10	0.47	0.20	0.19	0.04
Log Word Count	0.00	16.45	4.14	3.97	1.16
Readability	0.29	8.45	1.83	1.83	0.10
Attachment	0.00	1.00	0.12	0.00	0.32
Non-electronic Request	0.00	1.00	0.03	0.00	0.17
Inappropriate Request	0.00	0.64	0.09	0.08	0.03
Multiple Queries	0.00	1.00	0.42	0.35	0.42
Multiple Areas	0.23	1.00	0.89	0.91	0.09
Unrelated Queries	0.02	0.99	0.52	0.54	0.22
Information Type - Datum	0.00	0.96	0.10	0.04	0.15
Information Type - Data	0.00	1.00	0.69	0.76	0.24
Information Type - Database	0.00	0.91	0.11	0.07	0.12
Information Type - Document	0.00	0.96	0.15	0.08	0.16
Information Type - Multiple Documents	0.00	0.99	0.19	0.11	0.20
Information Theme - Activities	0.01	0.98	0.45	0.45	0.20
Information Theme - Budgets	0.00	1.00	0.17	0.09	0.21
Information Theme - Evaluation	0.00	0.99	0.21	0.15	0.18
Information Theme - Contracts	0.00	1.00	0.15	0.05	0.23
Information Theme - Personnel	0.00	1.00	0.34	0.24	0.29
Information Theme - Regulatory	0.00	0.98	0.10	0.06	0.12
Information Theme - Other	0.00	0.23	0.01	0.01	0.01

Table E.2: Summary statistics of the full dataset analysed in the models in Table 2 of the main paper.

F Analyses of Demand-Side Feedback Effects

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Muni. Lagged Avg. Log Days-to-Response	0.0872^{***}	0.0380^{***}								
	(0.0135)	(0.0111)								
Muni. Lagged Avg. Official Inexistencia			0.3326^{***}	0.0236						
			(0.0985)	(0.0387)						
Muni. Lagged Avg. True Inexistencia					0.2609^{***}	0.0914^{**}				
					(0.0689)	(0.0373)				
Muni. Lagged Avg. Uninformative Response							-0.1247^{***}	-0.0787^{***}		
							(0.0298)	(0.0250)		
Muni. Lagged Avg. Noncompliant Response									-0.0061	0.0095
									(0.0322)	(0.0173)
Num. obs.	1,145,164	1,145,164	1,145,164	1,145,164	1,144,822	1,144,822	1,144,822	1,144,822	1,137,705	1,137,705
Adj. R^2	0.0426	0.0991	0.0155	0.0951	0.0196	0.0966	0.0099	0.0978	0.0000	0.0946
FE: Municipality		X		X		X		X		X

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table F.3: Linear models of request expertise, for requests 2003-2016. Results correspond to the first two columns in Table 1 in the main manuscript, but replacing the time indicator with measures of average responsiveness to requests filed from each municipality in the previous year. Standard errors clustered by municipality.

G Additional Interaction Term Figures Based on Main Paper Tables 2 and 3

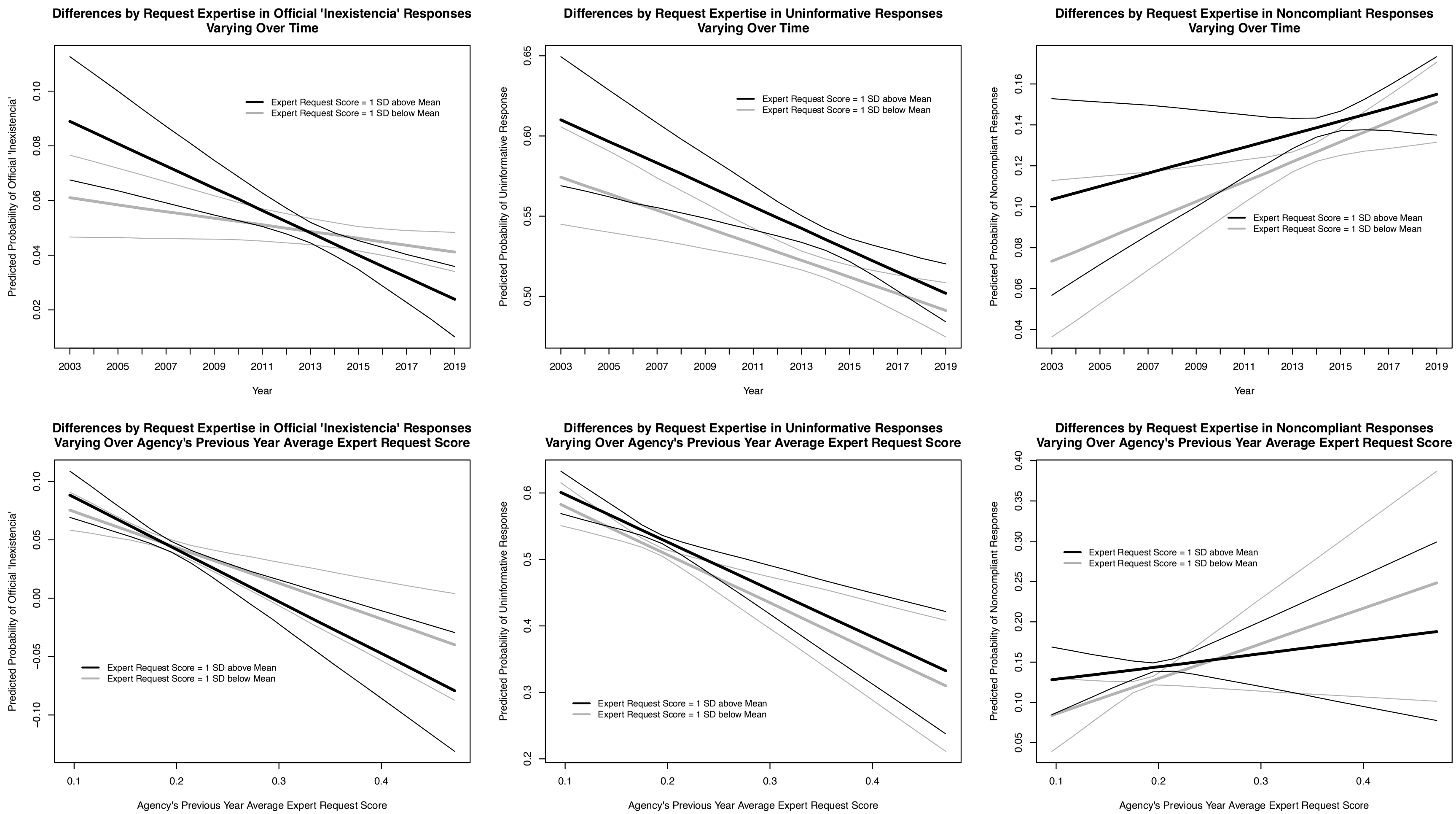


Figure G.4: Visualizations of key results from Models 2, 4, and 5 in Table 2 and Models 2, 4, and 5 in Table 3. These are based on the same comparisons of otherwise-similar hypothetical requests as in Figure 2 in the main paper.

H Nonlinear Interaction Test Figures

Figure H.1: Interflex Plots for Models 1-3 in Table 2 (Left Column) and Table 3 (Right Column)

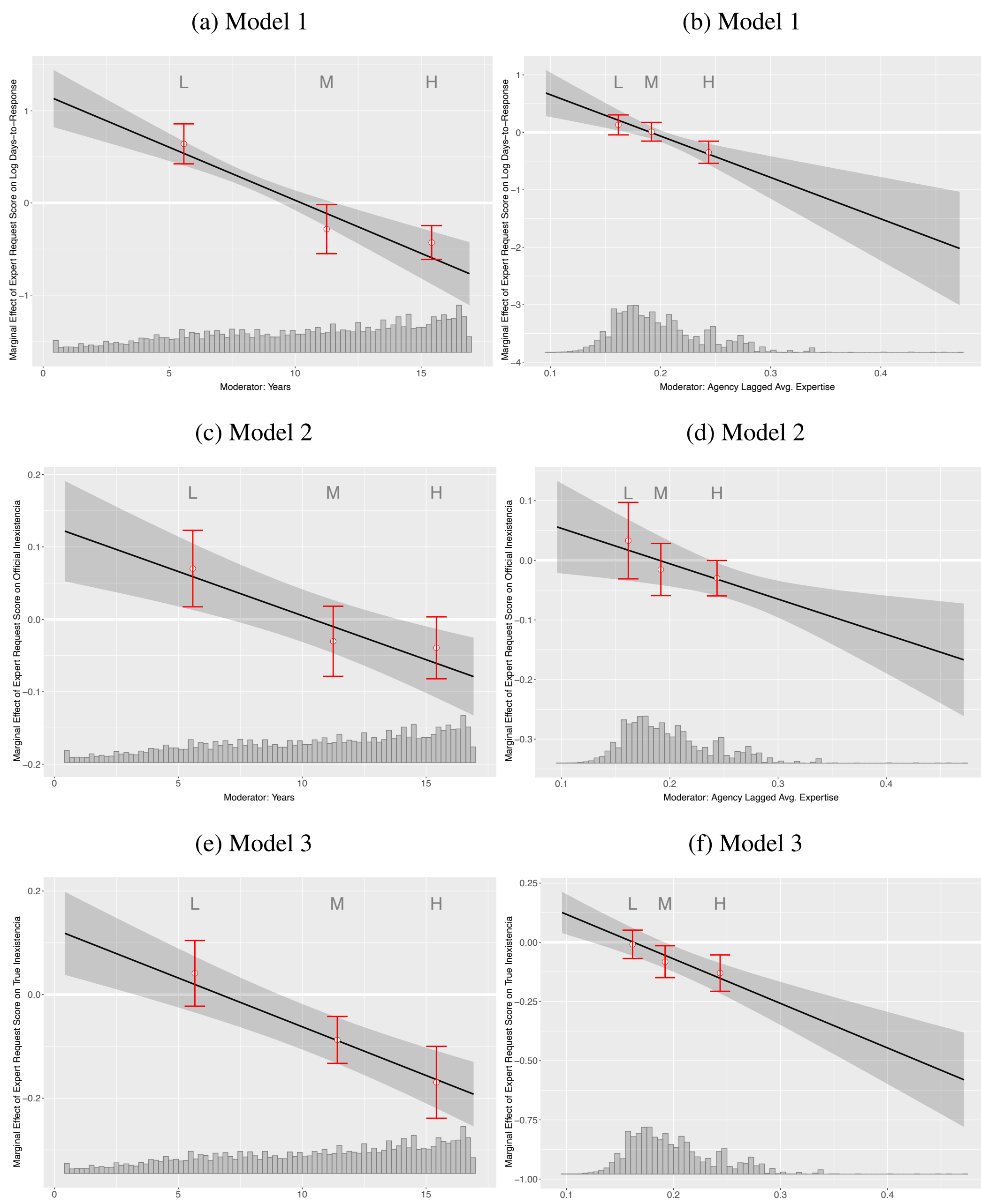
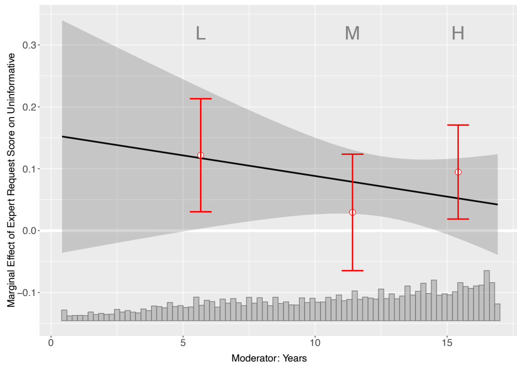
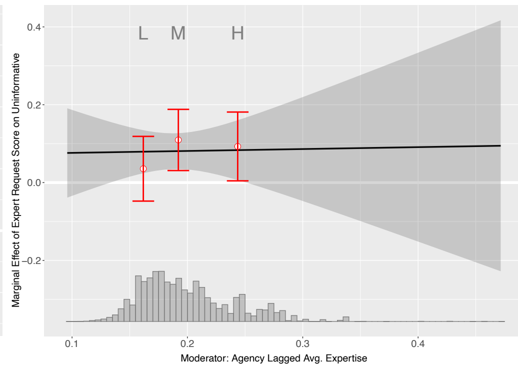


Figure H.2: Interflex Plots for Models 4-5 in Table 2 (Left Column) and Table 3 (Right Column)

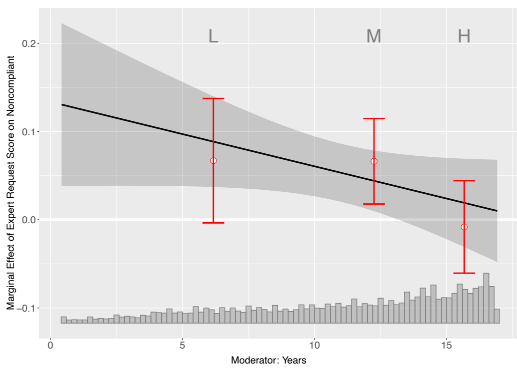
(a) Model 4



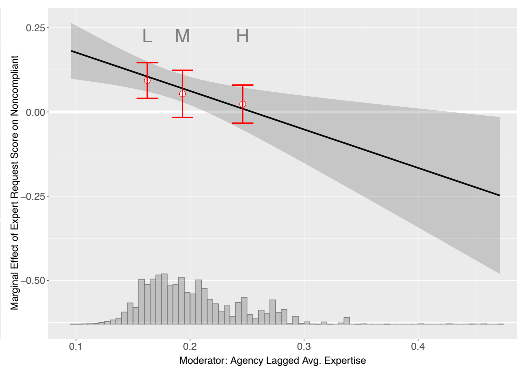
(b) Model 4



(c) Model 5



(d) Model 5



I Robustness Checks

Dependent Variable	Model 1 Log Days-to-Response	Model 2 Official Inexistencia	Model 3 True Inexistencia	Model 4 Uninformative	Model 5 Noncompliant
Expert Request Score (PCA version)	0.801^{***}	0.054^{***}	0.039	0.007	0.034
	(0.085)	(0.020)	(0.026)	(0.055)	(0.028)
Years	0.000	-0.003^{***}	0.006^{***}	-0.006^{***}	0.004^{**}
	(0.004)	(0.001)	(0.001)	(0.001)	(0.002)
Expert Request Score × Years	-0.071^{***}	-0.007^{***}	-0.010^{***}	-0.001	-0.003
	(0.011)	(0.002)	(0.002)	(0.004)	(0.002)
Adj. R^2	0.174	0.085	0.062	0.108	0.068
Num. obs.	1,610,997	1,611,289	1,537,385	1,537,385	924,920
Controls	Y	Y	Y	Y	Y
Agency Fixed Effects	Y	Y	Y	Y	Y

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors clustered by agency.

Table I.1: Robustness check with alternate version of Expert Request Score, using primary component of a principal-components analysis of nine individual measures of request language and specificity. Linear models of request-level poor responsiveness. All models include agency fixed effects and controls for the month-of-year, each request's logged word count, readability, medium, inclusion of an attachment, as well as predicted measures of request appropriateness, complexity, and theme.

Dependent Variable	Model 1 Log Days-to-Response	Model 2 Official Inexistencia	Model 3 True Inexistencia	Model 4 Uninformative	Model 5 Noncompliant
Expert Request Score (PCA version)	0.058	-0.017	-0.062^{***}	-0.002	0.011
	(0.035)	(0.012)	(0.018)	(0.015)	(0.015)
Agency Lagged Avg. Expertise (PCA version)	-0.189	-0.201^{***}	0.256^{***}	-0.424^{***}	0.130
	(0.246)	(0.052)	(0.079)	(0.090)	(0.133)
Expert Request Score $\times$ Agency Lagged Avg. Expertise	-2.908^{***}	-0.218^{**}	-0.621^{***}	0.050	-0.377^{**}
	(0.456)	(0.104)	(0.175)	(0.168)	(0.170)
Adj. R^2	0.174	0.085	0.059	0.107	0.066
Num. obs.	1,577,378	1,577,605	1,507,203	1,507,203	909,650
Controls	Y	Y	Y	Y	Y
Agency Fixed Effects	Y	Y	Y	Y	Y

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors clustered by agency.

Table I.2: Robustness check with alternate version of Expert Request Score, using primary component of a principal-components analysis of nine individual measures of request language and specificity. Agency Lagged Average Expertise also recalculated on the basis of this alternate version. Linear models of request-level poor responsiveness. All models include agency fixed effects and controls for month-of-year, each request's logged word count, readability, medium, inclusion of an attachment, as well as predicted measures of request appropriateness, complexity, and theme.

Dependent Variable	Model 1 Log Days-to-Response	Model 2 Official Inexistencia	Model 3 True Inexistencia	Model 4 Uninformative	Model 5 Noncompliant
Expert Request Score	1.170^{***}	0.128^{***}	0.125^{***}	0.183^{*}	0.146^{***}
	(0.158)	(0.034)	(0.039)	(0.094)	(0.046)
Years	0.023^{***}	-0.000	0.009^{***}	-0.004^{**}	0.006^{***}
	(0.004)	(0.001)	(0.001)	(0.002)	(0.002)
Expert Request Score $\times$ Years	-0.114^{***}	-0.012^{***}	-0.019^{***}	-0.009	-0.009^{**}
	(0.018)	(0.003)	(0.003)	(0.007)	(0.004)
Adj. R^2	0.189	0.088	0.069	0.114	0.075
Num. obs.	1,728,620	1,728,922	1,646,272	1,646,272	1,000,529
Controls	Y	Y	Y	Y	Y
Agency Fixed Effects	Y	Y	Y	Y	Y

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors clustered by agency.

Table I.3: Robustness check with larger sample including all agencies that appear in the data. Linear models of request-level poor responsiveness. All models include agency fixed effects and controls for the month-of-year, each request's logged word count, readability, medium, inclusion of an attachment, as well as predicted measures of request appropriateness, complexity, and theme.

Dependent Variable	Model 1 Log Days-to-Response	Model 2 Official Inexistencia	Model 3 True Inexistencia	Model 4 Uninformative	Model 5 Noncompliant
Expert Request Score	1.214^{***}	0.096^{*}	0.259^{***}	0.053	0.262^{***}
	(0.346)	(0.055)	(0.069)	(0.100)	(0.069)
Agency Lagged Avg. Expertise	0.976^{**}	-0.239^{***}	0.730^{***}	-0.631^{***}	0.480^{**}
	(0.427)	(0.072)	(0.159)	(0.177)	(0.237)
Expert Request Score × Agency Lagged Avg. Expertise	-6.482^{***}	-0.521^{***}	-1.689^{***}	0.148	-1.025^{***}
	(1.722)	(0.201)	(0.330)	(0.522)	(0.358)
Adj. R^2	0.188	0.088	0.066	0.113	0.073
Num. obs.	$1,686,212$	$1,686,445$	$1,608,148$	$1,608,148$	$980,072$
Controls	Y	Y	Y	Y	Y
Agency Fixed Effects	Y	Y	Y	Y	Y

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors clustered by agency.

Table I.4: Robustness check with larger sample including all agencies that appear in the data. Linear models of request-level poor responsiveness. All models include agency fixed effects and controls for the month-of-year, each request's logged word count, readability, medium, inclusion of an attachment, as well as predicted measures of request appropriateness, complexity, and theme.

Dependent Variable	Model 1 Log Days-to-Response	Model 2 True Inexistencia	Model 3 Uninformative	Model 4 Noncompliant
Expert Request Score (handcoded)	0.222^{**}	0.075	0.183^{***}	0.173^{***}
	(0.111)	(0.064)	(0.063)	(0.058)
Years	0.027^{***}	0.010^{***}	0.001	0.002
	(0.007)	(0.002)	(0.003)	(0.003)
Expert Request Score (handcoded) × Years	-0.028^{**}	-0.009	-0.016^{**}	-0.012^{*}
	(0.013)	(0.008)	(0.007)	(0.007)
Adj. R^2	0.065	0.031	0.055	0.056
Num. obs.	$4,592$	$4,586$	$4,527$	$4,586$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors clustered by agency.

Table I.5: Robustness checks using hand-coded sample only. All classifier-predicted measures are here instead replaced with alternate versions based on original hand-coded request and response features. Linear models of request-level poor responsiveness.

Dependent Variable	Model 1 Log Days-to-Response	Model 2 True Inexistencia	Model 3 Uninformative	Model 4 Noncompliant
Expert Request Score (handcoded)	0.720^{***}	0.052	0.127	0.170^{*}
	(0.186)	(0.138)	(0.116)	(0.101)
Agency Lagged Avg. Expertise	2.127^{***}	-0.121	0.067	0.316
	(0.699)	(0.264)	(0.194)	(0.197)
Expert Request Score (handcoded) × Agency Lagged Avg. Expertise	-3.674^{***}	-0.191	-0.354	-0.449
	(0.810)	(0.726)	(0.564)	(0.456)
Adj. R^2	0.064	0.026	0.057	0.058
Num. obs.	$4,464$	$4,458$	$4,403$	$4,458$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors clustered by agency.

Table I.6: Robustness checks using hand-coded sample only. All classifier-predicted measures are here instead replaced with alternate versions based on original hand-coded request and response features; with the exception of Agency Lagged Avg. Expertise. Linear models of request-level poor responsiveness.

Dependent Variable	Model 1 Log Days-to-Response	Model 2 Official Inexistencia	Model 3 True Inexistencia	Model 4 Uninformative	Model 5 Noncompliant
Expert Request Score	1.058^{***}	0.104^{***}	0.138^{***}	0.120	0.122^{**}
	(0.150)	(0.034)	(0.042)	(0.088)	(0.051)
Years	0.020^{**}	-0.002	0.008^{***}	-0.011^{***}	0.001
	(0.009)	(0.002)	(0.002)	(0.003)	(0.005)
Expert Request Score × Years	-0.103^{***}	-0.010^{***}	-0.019^{***}	-0.004	-0.007
	(0.017)	(0.003)	(0.004)	(0.007)	(0.004)
Presidency: Fox	-0.015	-0.003	-0.010	-0.075^{***}	-0.048
	(0.067)	(0.021)	(0.016)	(0.025)	(0.041)
Presidency: Calderon	0.072	0.005	-0.011	-0.056^{***}	-0.025
	(0.056)	(0.015)	(0.012)	(0.016)	(0.028)
Presidency: EPN	0.007	-0.002	-0.022^{***}	-0.011	0.009
	(0.022)	(0.009)	(0.008)	(0.013)	(0.019)
Policy Period 2	0.120^{***}	0.033^{***}	0.011	0.034^{**}	-0.005
	(0.032)	(0.008)	(0.009)	(0.017)	(0.013)
Policy Period 3	0.117^{**}	0.044^{***}	0.025^{*}	0.046^{**}	0.009
	(0.049)	(0.010)	(0.014)	(0.018)	(0.018)
Policy Period 4	0.109^{**}	0.038^{***}	0.029^{**}	0.027	0.009
	(0.051)	(0.012)	(0.014)	(0.019)	(0.021)
Adj. R^2	0.178	0.087	0.062	0.109	0.069
Num. obs.	1,610,997	1,611,289	1,537,385	1,537,385	924,920
Controls	Y	Y	Y	Y	Y
Agency Fixed Effects	Y	Y	Y	Y	Y

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors clustered by agency.

Table I.7: Robustness check including additional controls for time-periods of each presidency, and of legal and constitutional changes in July 2007, February 2014, and May 2015. Linear models of request-level poor responsiveness. All models include agency fixed effects and controls for the month-of-year, each request's logged word count, readability, medium, inclusion of an attachment, as well as predicted measures of request appropriateness, complexity, and theme.

Dependent Variable	Model 1 Log Days-to-Response	Model 2 Official Inexistencia	Model 3 True Inexistencia	Model 4 Uninformative	Model 5 Noncompliant
Expert Request Score	1.366^{***}	0.138^{**}	0.241^{***}	0.113	0.223^{***}
	(0.422)	(0.054)	(0.067)	(0.118)	(0.077)
Agency Lagged Avg. Expertise	0.954	-0.070	0.307^{**}	-0.430^{**}	0.045
	(0.635)	(0.134)	(0.147)	(0.196)	(0.185)
Expert Request Score $\times$ Agency Lagged Avg. Expertise	-7.106^{***}	-0.704^{***}	-1.555^{***}	-0.172	-0.856^{**}
	(2.022)	(0.203)	(0.311)	(0.596)	(0.371)
Presidency: Fox	-0.011	0.027^{*}	-0.035^{***}	0.004	-0.040^{**}
	(0.049)	(0.015)	(0.011)	(0.015)	(0.016)
Presidency: Calderon	0.068	0.026^{**}	-0.031^{***}	0.004	-0.022
	(0.049)	(0.012)	(0.009)	(0.012)	(0.014)
Presidency: EPN	0.011	0.007	-0.029^{***}	0.014	0.011
	(0.022)	(0.008)	(0.008)	(0.011)	(0.012)
Policy Period 2	0.141^{***}	0.023^{***}	0.026^{***}	0.002	-0.004
	(0.039)	(0.009)	(0.007)	(0.017)	(0.015)
Policy Period 3	0.143^{***}	0.032^{***}	0.045^{***}	0.005	0.011
	(0.047)	(0.011)	(0.012)	(0.018)	(0.014)
Policy Period 4	0.125^{**}	0.013	0.057^{***}	-0.046^{**}	0.008
	(0.050)	(0.013)	(0.009)	(0.018)	(0.017)
Adj. R^2	0.177	0.088	0.062	0.109	0.069
Num. obs.	1,577,378	1,577,605	1,507,203	1,507,203	909,650
Controls	Y	Y	Y	Y	Y
Agency Fixed Effects	Y	Y	Y	Y	Y

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors clustered by agency.

Table I.8: Robustness check including additional controls for time-periods of each presidency, and of legal and constitutional changes in July 2007, February 2014, and May 2015. Linear models of request-level poor responsiveness. All models include agency fixed effects and controls for the month-of-year, each request's logged word count, readability, medium, inclusion of an attachment, as well as predicted measures of request appropriateness, complexity, and theme.

J Qualitative Research Appendix

Qualitative field research consisted mainly of semi-structured interviews with representatives of Mexican civil society organizations (CSOs) and staff employed in federal government agencies' Transparency Units (Unidades de Transparencia). Qualitative findings reported in the paper draw on interviews with 14 organizations (listed in Table J.1), seven government ministries or agencies, and three individuals, previously affiliated with CSOs working in areas of transparency and/or INAI (Juan Pablo Guerrero, Ana Joaquina Ruiz, Maylí Sepúlveda). We do not provide any identifying information about Transparency Unit staff because these participants were promised anonymity.

Interviews with CSOs were conducted over Zoom by two of the authors between September 2020 and January 2021. Organizations were selected to represent variation in sector (education, environment, gender rights, health, human rights). Within each sector, we interviewed at least one "elite" organization—with a high level of human and financial resources, typically based in Mexico City—and one "non-elite" organization with lower resources, based outside the capital. We also interviewed representatives of several "broker" organizations, which work in multiple policy areas and specialize in navigating institutions of social accountability. Selection of cases was aided by an institutional partner, Fundar, a leading think tank on transparency and accountability in Mexico as well as snowball sampling.We provide names of organizations in this paper, but not the names of the persons interviewed (although all consented to be identified). Interviews with CSOs were audio recorded and transcribed using Amazon Web Services. We then coded transcriptions for theoretically relevant variables using NVivo. Anonymized transcripts are available upon request.

Organization	**Sector**	**Profile**
Mejora Tu Escuela	Education	Elite
Centro Interdisciplinario de Desarrollo Alternativo	Environment	Non-Elite
Mujeres de Caltongo	Environment	Non-Elite
Oceana	Environment	Elite
Equis Justicia para Mujeres	Gender Rights	Elite
Instituto de Liderazgo Simone de Beauvoir	Gender Rights	Non-Elite
Derechohabientes Viviendo con VIH	Health	Non-Elite
Nosotrxs	Health	Elite
Mentte CEDAT	Human Rights	Non-Elite
Poder	Human Rights	Elite
Artículo 19	Broker	Elite
Controla Tu Gobierno	Broker	Elite
Fundar	Broker	Elite
SocialTIC	Broker	Elite

Table J.1: Civil society organizations participating in field research

Interviews with Transparency Unit staff were conducted in person by one of the authors and a research assistant in March 2017. Seven entities were selected from among the twenty-five ministries and agencies with the highest total request volumes overall at that time, but with attention to diversity across policy areas and functions (e.g., four of the seven were cabinet-level ministries). Interviews were conducted in person at each entity's offices in Mexico City. Some interviews were audio recorded and transcribed, while for others the author simply took written notes.

J.1 Ethical Considerations

Prior to conducting field research with CSOs, we received approval from [REDACTED] Institutional Review Board for the Protection of Human Subjects (Protocol No: HR-3662, "How Do Accountability Institutions Change Over Time? Big Data and Access to Information in Mexico," 8/14/2020). Prior to conducting field research with Transparency Unit staff, we received Institutional Review Board approval from [REDACTED] (Protocol No: 00005773, "Implementing Access to Information in Mexican Government Agencies," 2/21/2017). Both of these approval processes determined protocols to be Exempt from Institution Review Board scrutiny. Nonetheless, we took pains to uphold the highest standards of ethics. We summarize our practices in response to excerpted principles from APSA's Principles and Guidance for Human Subjects Research.

Power: *"When designing and conducting research, political scientists should be aware of power differentials between researcher and researched, and the ways in which such power differentials can affect the voluntariness of consent and the evaluation of risk and benefit."*

Most participants in our study do not classify as low-power or vulnerable populations, as all interviewees were either mid-level federal government officials or leaders of civil society organizations. The possible exceptions are leaders of non-elite civil society organizations who in some cases came from vulnerable (i.e. low-income, rural, and or indigenous) communities. In such cases, we took extra efforts to explain the goals of the research and potential uses for information gathered in interviews and to insist that participation in our study was entirely voluntary. All organization leaders expressed eagerness to relate their organizations' experiences and to draw attention to shortcomings in state institutions of accountability.

Consent: *"Political science researchers should generally seek informed consent from individuals who are directly engaged by the research process, especially if research involves more than minimal risk of harm or if it is plausible to expect that engaged individuals would withhold consent if consent were sought."*

In compliance with our IRB proposals, we requested (and were granted) informed written consent by CSO participants. For interviews with transparency unit staff, interviewees were given a written consent document, and read a summary of this, but were not asked for signatures as these would conflict with the assurances that they would not be individually identified. We have no reasons to believe that risks posed to participants have changed significantly since we began our project and thus have not returned to participants for continuing consent to use information gathered in interviews.

Deception: *"Political science researchers should carefully consider any use of deception and the ways in which deception can conflict with participant autonomy."*

Our research did not include deception or covert methods.

Harm and Trauma: *Political science researchers should consider the harms associated with their research.*

We agree with IRB appraisals that our human subjects research posed low risk of harm to participants. Potential harm to civil society participants may result from reprisals from government officials. Given that participants were engaged in public activism campaigns, they were aware of such risks. Potential harm to Transparency Unit personnel may result from reprisals from agency superiors. To minimize such risks—and to encourage participants to speak openly—we maintain the names of these personnel and their agencies confidential.

Confidentiality: *Political science researchers should generally keep the identities of research* *participants confidential; when circumstances require, researchers should adopt the higher standard of ensuring anonymity.*

As stated above, we promised to (and sustain) the confidentiality of government agents that participated in research owing to heightened risk of professional reprisals. We offered to sustain confidentiality of civil society participants, but none requested it. In most cases, civil society actors were eager to draw attention to their causes and had significant experience publicizing their activities and demands to the media. We thus include the names of CSOs, but not the names of their representatives whom we interviewed, given that we see little analytical benefit to be gained from revealing the latter.

Impact: *Political science researchers conducting studies on political processes should consider the broader social impacts of the research process as well as the impact on the experience of individuals directly engaged by the research. In general, political science researchers should not compromise the integrity of political processes for research purposes without the consent of individuals that are directly engaged by the research process.*

This academic research project was accompanied by a significant knowledge exchange component. Through this we sought to share methodological tools for evaluating the performance of Mexico's access-to-information system and recommendations to civil society actors and government personnel to sustain optimal functioning of this system. We have taken pains to emphasize that our recommendations are specific of the research team and do not reflect the opinions of any research participants. At the onset of these activities we presented our research plan in listening sessions with officials from Mexico's Access-to-Information Institute (INAI) and civil society leaders and engaged in ongoing communication with these figures throughout the project. On the invitation of INAI, we publicly presented key results of our study on [DATE REDACTED].

Laws, Regulations, and Prospective Review: *Political science researchers should be aware of relevant laws and regulations governing their research related activities.*

All government data that we used for this project are publicly available under Mexican law, including principally citizen requests for public information and corresponding responses from government agencies. (Our dataset does not include requests for Personal Information, the texts of which are not made publicly available.) Citizens submitting information requests are advised not to include identifying information in the text of their requests. Nonetheless, on rare occasions, citizens violate these norms. For this reason, we do not include information request texts themselves in our publicly available replication data.

Shared Responsibility: *The responsibility to promote ethical research goes beyond the individual researcher or research team.*

All members of our research team participated in drafting and upholding human subjects protocols. To avoid any potential violation of confidentiality of citizens submitting information requests, we refrain from publicizing information request texts themselves.

K Generating Training Data: Hand-coding

We drew a random sample of 6,000 requests (and their corresponding responses) to generate our gold standard hand-coded data. Our random sample was composed of a simple random sample of 4,000 requests for which the official response was that the information was provided electronically (61.8% of the total), and thus for which response documents existed and could be analyzed further. In addition, we drew an over-sample of 2,000 requests pertaining to the forty agencies that received the highest volume of requests over the 2003-2015 period (50 additional sampled requests per agency). This oversampling was intended to ensure that our training data were appropriate in particular for any future analyses that focused predominantly on these agencies.

To code variables for this sample of request-response texts, a team of six human-coders completed a questionnaire related to the 6,000 sampled requests and their responses mentioned above. All members of this coding team were Mexican nationals and were fully fluent in Spanish. The team manager (an author) had prior work experience with Mexico's Access-to-Information Institute. All coders furthermore were graduates of CIDE—a well-recognized center for research and higher education in Mexico with a particular focus on economics and the social sciences.

All human-coding of our request and response texts was completed in Spanish. A bilingual version of the questionnaire that the coders filled out in an online interface is available upon request. To ensure high inter-coder reliability, we conducted extensive training led by the two authors on the present paper. This training was done in person in Mexico City and involved twelve hours of preliminary training sessions as well as three calibration meetings after initial waves of test codings were conducted.[1] An initial pilot round of coding was excluded from further use. Of the remaining sampled requests/responses, 543 were coded by multiple coders (between two and four each) to enable assessment of inter-coder reliability. Not all of the 6,000 initially sampled requests were ultimately hand-coded, due to time and resource constraints (but the order in which they were coded was random). The final hand-coded sample to be used as training data for machine learning models thus comprised 4,934 requests and their associated responses.

Our assessment of inter-coder reliability for the 543 multiply coded request-response pairings suggests that a majority of our request and response features are valid as based upon the consistency of our human-codings. In Figures K.1a and K.1b, we report three traditional measures of inter-coder reliability: the intra-class correlation (ICC), Krippendorff's α and Kendall's Coefficient of Concordance (W) for our request and response variables, respectively. We also show the average of these three reliability measures via the circles that appear on each plot.

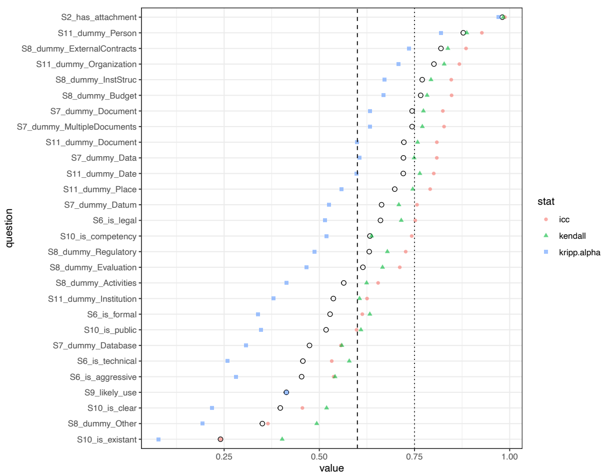
As can be seen in these figures, many of our individual request and response features demonstrate a high degree of inter-coder reliability based off of our questionnaire items, while others less so. Both plots report generally accepted[2] thresholds for good-to-substantial and substantial-to-excellent inter-coder reliability of 0.6 and 0.75, respectively via a pair of vertical dashed lines. For the request variables in Figure K.1a, 17 out of 28 (61%) lie above the good-to-substantial threshold of 0.6 based upon average inter-coder reliability values, with several such variables exhibiting substantial-to-excellent inter-coder reliability of 0.75 or higher. Only seven out of our 28 request variables (25%) then fall below the value of 0.5 in their average levels of inter-coder reliability—a threshold that could conservatively be seen as a reasonable cutoff for moderate-to-good inter-coder reliability.[3] Importantly, note that a majority of these less reliably coded variables are not central to our analyses. For example, of the nine variables that were used to construct our key independent variable measure of request expertise, six (67%) had average inter-coder reliability values over the 0.6 threshold and eight (89%) had average inter-coder reliability values over the 0.5 threshold.

Our response variables' inter-coder reliability scores in Figure K.1b indicate even more relia-

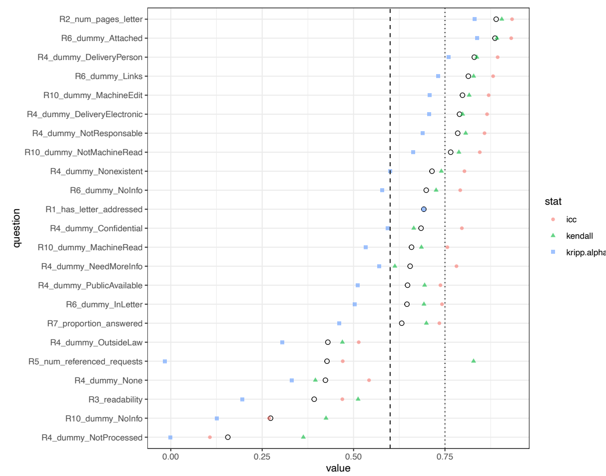
1. Coder training and human-coding were each undertaken during a single 10-week period in 2019.
2. E.g., Steiner et al. (2004), Tang et al. (2015), Bächtiger and Hangartner (2010).
3. See, e.g., Steiner et al. (2004), Tang et al. (2015), and Bächtiger and Hangartner (2010, Note 5).

Figure K.1: ICR measures. Vertical dashed lines represent a cutoff of .6 and vertical dotted lines represent a cutoff of .75.

(a) Request Variables



(b) Response Variables



bility and coding consistency than was the case for our request variables. In total, 17 out of our 23 response variables (i.e., 74%) exhibited inter-coder reliability averages that were greater than 0.6 (i.e., our good-to-substantial cutoff for inter-coder reliability). Eight of these 17 measures (i.e., 35% of all response variables) furthermore exhibited an average inter-coder reliability score greater than 0.75—a level of inter-coder reliability that is commonly viewed as substantial-to-excellent (Steiner et al. 2004; Tang et al. 2015; Bächtiger and Hangartner 2010, Note 5). Finally, we emphasize that *none* of the response variables that fall below the 0.6 threshold have been employed in the present analysis. Altogether this implies that our paper's response-based human-codings (and the measures subsequently derived from them) are highly reliable.

With these levels of inter-coder reliability established, we define a method for combining our 543 jointly coded request and response variables into single scores for use in our subsequent machine learning steps, alongside the variable scores produced by single coders only. Our method for combining these jointly coded scores is as follows. For each observation and each variable that was jointly human-coded: if the majority of the human-coders who coded the data point agreed on a code, that majority-code was assigned for subsequent analysis. If the coders disagreed, we broke the tie based on our ranked assessment of the coders' ability and experience, with input from the training session and coding manager. After combining these jointly coded request and response variables, our full sample of human-coded requests and responses was passed on to a series of supervised classifiers, so as to use these human-coded variables to classify all information request texts and response texts in our sample. This machine learning methodology is described next.

L Machine Learning Models

We now describe our supervised machine learning approach. This was separately developed for our request- and response-level codings. We begin with a discussion of the methodology used for our request-level variables, followed by our response-level machine learning efforts.

L.1 Requests

Recall that our main paper's analyses relied upon multiple request-level independent and control variables. As discussed therein, many of these variables were derived from our individual ATI request texts using supervised machine learning. Examples include the level of expertise exhibited by each requester (our primary independent variable) and control variable measures of the appropriateness, complexity, and theme of each information request. We now outline how we generated these request-level variables using supervised machine learning methods.

Despite recent advances in supervised machine learning techniques applied to political texts, a majority of these techniques are designed to predict a only single variable for each individual textual document. When faced with multiple target variables to predict from the same text, researchers tend to ignore these target variables' levels of interdependence during supervised classification. In this context, substantial gains in predictive performance may be obtained by treating target variables as interdependent—leveraging each variable's supervised predictions as features during the supervised classification of all other variables (Erlich, Dantas, Bagozzi, Berliner, and Palmer-Rubin 2021). This framework is known as multi-label classification.

In line with this multi-label framework, our own analysis requires that we use supervised machine learning methods to generate multiple request-level variables for each individual ATI request. To classify our requests based on our hand-coded request sample in this manner, we therefore explored the benefits of the multi-label supervised machine learning framework described immediately above. This framework leverages the inter-label relationships among our request variables so as to potentially improve the accuracy of predicting our non-mutually exclusive traits. In other words, our predictive optimization task under this multi-label framework entails the joint classification of all proper label sets for our unlabeled requests, instead of independent label prediction.

To this end, we first sought to verify the value-added of multi-label prediction in our particular application. Here we first estimated a selection of multi-label and non-multi-label supervised classifiers on 80-20 (training-test) splits of our hand-coded requests for 26 distinct hand-labeled request traits. The best performing binary classifier that did not take into consideration inter-label dependence was an approach known as binary relevance (BR). The best multi-label classification technique was found to be an approach known as ensemble classifier chain (ECC). Both approaches are defined in detail in Erlich et al. (2021). We next present a comparison of these two best performing techniques for our current application.

Our BR and ECC comparison focuses on the classification of our request trait labels within our hand-labeled sample of 4,934 requests. We utilize 10-fold cross-validation alongside the 80-20 splits mentioned above. The hand-coded traits that we consider encompass the request variables highlighted in the previous hand-coding and inter-coder reliability sections of this appendix. Where applicable, polytomous measures were dichotomized aside from our hand-coded measure of each requester's request area,[4] which we maintained in its polytomous form.

The above features correspond to the "outcome variables" that our supervised methods seek to predict. For the independent variables that we use in predicting these outcome variables, we primarily leverage the vectorized response texts themselves. In addition to these vectorized response texts, we then further optimize these models by providing our ECC and BR classifiers with the following additional request level predictors: 1) agency, 2) year, 3) number of attached documents, 4) sector, 5) state, and 6) municipality. Across these varied features, we find that the vectorized

4. I.e., if the request was academic/scholarly, commercial, monitoring-focused, personal, or impossible to say.

text provides by far the best predictions.[5] This suggests that the additional predictors mentioned above provide only a negligible contribution to our request labels' predictions.

The overall results of the BR and ECC models are presented in Table L.1. Each set of reported results represents the best performing (BR or ECC) model across a range of candidate base classifiers.[6] Given our multi-label context, we consider several specialized classification metrics to compare our two supervised machine learning approaches, as opposed to traditional metrics such as precision and recall (Erlich et al. 2021). As seen in Table L.1, the ECC model performed better for 3 out of 5 metrics: subset accuracy, F1-micro, and F1-macro. Moreover, the improvements of ECC over BR in at least two of these cases (i.e., subset accuracy and F1-macro) were substantial. This is in contrast to the instances of superior performance by BR (i.e., hamming loss and ranking loss), which each suggest relatively marginal improvements over ECC.

Algorithm	**Subset Accuracy**	**Hamming Loss**	**Ranking Loss**	**F1-micro**	**F1-macro**
Ensemble CC (ECC)	8.22 % ± 0.75	11.89%± 0.17	0.076 ± 0.002	77.59 ± 0.27	45.20 ± 0.76
Binary Relevance (BR)	4.60% ± 0.66	11.42% ± 0.19	0.071± 0.002	76.84± 0.31	39.84± 0.64

Table L.1: Results

We accordingly conclude that ECC—and hence multi-label prediction—is an optimal approach for classifying our request traits. Our final step in generating these classified request features leads us to re-train our ECC framework on our full set of hand-coded requests and labels. We then use this full ECC model to predict our request trait labels for our full sample of ATI requests. Our final request measures thus reflect these ECC predictions of each request label across all ATI requests.[7]

L.2 Responses

To classify ATI responses, we turn to our hand-codings of individual response texts. These hand-coded responses correspond to the agency responses that were made to same 4,934 request sample described above. Note, however, that we do not use any information about the requests themselves—either in terms of vectorized request texts or request labels—for our response classifications. Rather, and using our response hand-codings and their underlying response texts, we sought to predict our relevant response features for the entire universe of responses. In this case, our primary response features correspond to our hand-codings of the actual responses and quality of information provided to each requester. Recall that these dependent variable inputs denote our assessments of whether an officially designated positive response was found to be misleading in the sense that it actually included a denial, or if information was actually provided in that response.

5. In ECC, note that the "previous labels" are also used as features by design. The previous labels are not explicit features, they are consequence of the chain of binary classifiers that are employed in the classifier chain model. We employed the logistic regression as base classifier for all multi-label methods.

6. In particular, we considered logistic regression, gradient boosting, and random forests as base classifiers. Logistic regression performed best as a base classifier for both BR and ECC.

7. All machine learning-based request-side variables in our analyses are in raw probability form rather than in dichotomized-probability form. This means that we must use our ECC predictions—as opposed to our actual hand-labels—for our approximately 5,000 labeled ATI requests within these variable measures, in order to ensure a consistent measurement scale across all observations for each request variable within our primary analyses.

For the supervised classification of our response features, we ultimately use binary classification, not multi-label classification. One reason for our doing so pertains to the more limited number of response features in this supervised classification context, when compared to our request traits application. As mentioned above, the efficacy of multi-label approaches to classify political text-as-data was evaluated in different settings in Erlich et al. (2021). In general, when features are abundant or more informative, the benefits of using multi-label models become less salient in comparison to the performance of individual predictions for each label (i.e., BR).

Additionally, since multi-label methods have higher computational cost than BR, the trade-off between performance and complexity/cost becomes less favorable if the performance gain from multi-label classification is minimal. Altogether, these insights lead us to favor BR in the context of our response predictions on theoretical grounds. That being said, we still directly compare both classification approaches for response label prediction below. Within these experiments, we found that the multi-label approaches considered did not yield a substantial enough increase in performance to justify the added complexity that they introduced into our predictive framework.

To this end, there are at least two additional reasons for ECC's relatively weak performance in this context. First, the composition of responses are more standardized than that of requests, owing to the larger heterogeneity in request text authors relative to response text authors. This increases our ability to predict response features independently. In other words, since Mexico's ATI responses are written by officials, not the general public as in the case of requests, the word distribution is narrower. Second, for our response texts, the additional features that were **not** predictive in the request case—such as the number of documents and the length of the response—were predictive in the response case. This is discussed in more detail below, and suggests that the relatively higher extant gains in available predictive features for our response texts likely ensures a more marginal multi-label contribution during prediction.

To more concretely present and assess the points made above, we now compare the predictive accuracy of our favored multi-label approach (i.e., ECC) to that of several BR approaches. These predictive evaluations follow a comparable structure to those conducted for our request evaluations presented above in terms of training-test splits, cross-validation, and evaluation sample. These details are hence not repeated here. In this response case, note that we also again consider the same potential features as in the request case in addition to the vectorized text: 1) agency, 2) year, 3) number of attached documents, 4) sector, 5) state, and 6) municipality. In this case, and in addition to the vectorized text, we found that the number of documents and their length had substantial predictive power. As noted above, this was not the case for our request text evaluations.

Our response classification experiments more specifically compare the results using our preferred multi-label classification method (ECC) to four different binary classification techniques: random forest, logistic regression and `XGBoost`.[8] The results from these comparisons are reported in Tables L.2 and L.3. Based upon Table L.2, we find that `XGBoost` performs best in classifying our relevant hand-coded response characteristics. Indeed, this model often produces competitive results without fine-tuning, especially in tabular data (Shwartz-Ziv and Armon 2021). In Table L.3, we likewise find `XGBoost` to be the best performing model using our multi-label evaluation criteria. Together these results suggest both that (1) BR-based approaches are superior to multi-label methods such as ECC for response classification and that (2) `XGBoost` is particularly optimal for our current response prediction tasks among all BR approaches considered. We

8. For details on `XGBoost`, see Chen and Guestrin (2016).

accordingly re-train our `XGBoost` classifier upon our full set of hand-coded response texts and labels, and use this classifier to in-turn predict all unlabeled response texts for each of our relevant response traits. The dependent variables utilized in the main manuscript are accordingly based upon supervised machine codings of response characteristics using `XGBoost` in a BR-framework.

Algorithm	**Average Accuracy**	**Average F1**
ECC	81.39	53.13
XGBoost	**82.64**	**66.00**
Logistic Regression	81.57	52.61
Random Forest	82.28	59.78

Table L.2: Standard Classification Statistics for Response Predictions

Algorithm	**Subset Accuracy**	**Hamming Loss**	**Ranking Loss**	**F1-micro**	**F1-macro**
ECC	58.57 %	18.69%	0.010	79.23	58.57
XGBoost	**58.89**%	**16.94**%	**0.008**	**81.23**	**66.97**
Logistic Regression	56.60%	18.42%	0.009	79.28	52.60
Random Forest	58.24%	17.80%	0.010	80.04	59.34

Table L.3: Multi-label Classification Statistics for Response Predictions

References

Bächtiger, André, and Dominik Hangartner. 2010. "When Deliberative Theory Meets Empirical Political Science: Theoretical and Methodological Challenges in Political Deliberation." *Political Studies* 58:609–629.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining,* 785–794.

Erlich, Aaron, Stefano G Dantas, Benjamin E Bagozzi, Daniel Berliner, and Brian Palmer-Rubin. 2021. "Multi-Label Prediction for Political Text-as-Data." *Political Analysis* 30 (4): 463–480.

Shwartz-Ziv, Ravid, and Amitai Armon. 2021. "Tabular Data: Deep Learning is Not All You Need." In *8th ICML Workshop on Automated Machine Learning (AutoML).*

Steiner, Jürg, André Bächtiger, Markus Spörndli, and Marco R. Steenbergen. 2004. *Deliberative Politics in Action: Analysing Parliamentary Discourse.* Cambridge University Press.

Tang, Wan, Jun Hu, Hui Zhang, Pan Wu, and Hua He. 2015. "Kappa Coefficient: A Popular Measure of Rater Agreement." *Shanghai Arch Psychiatry* 27 (1).