

Institutional Sources of Legitimate Authority:

An Experimental Investigation*

Running Head: Institutions and Legitimacy

Keywords: institutions, legitimacy, authority, public goods, experiment, incentives, compensation, transparency, political psychology

Eric S. Dickson

Associate Professor

New York University

Department of Politics

19 West 4th Street, 2nd Floor

New York, NY 10012

eric.dickson@nyu.edu

Sanford C. Gordon

Professor

New York University

Department of Politics

19 West 4th Street, 2nd Floor

New York, NY 10012

sanford.gordon@nyu.edu

Gregory A. Huber

Professor

Yale University

Department of Political Science and ISPS

77 Prospect Street

PO Box 208209

New Haven, CT 06520-8209

gregory.huber@yale.edu

*Earlier versions of this paper were presented at the NYU Law School's Law and Economics Workshop; the Seminar in Law and Politics at Harvard, the University of Pittsburgh Methods Seminar; the University of Texas at Dallas Economics Department Seminar; the NYU Psychology Department Seminar; the NYU-CESS Experimental Economics Seminar; the West Coast Experiments Conference; the Social Dilemmas Conference; the NYU-LSE Conference on Political Economy; the Political Economy of Public Law Conference; and the European Political Science Association, Economic Science Association, International Society of Political Psychology, and Midwest Political Science Association Annual Meetings, where the authors

Abstract

Unelected officials with coercive powers (e.g., police, prosecutors, bureaucrats) vary markedly in the extent to which citizens view their actions as legitimate. We explore the institutional determinants of legitimate authority in the context of a public goods laboratory experiment. In the experiment, an “authority” can target one “citizen” for punishment following citizen contribution choices. Untargeted citizens can then choose to help or hinder the authority. This latter choice may be interpreted as a behavioral measure of the authority’s legitimacy. We find that legitimacy is affected by how authorities are compensated, the transparency with which their decisions are observed, and an interaction between these. When transparency is high, citizens are more willing to assist authorities who receive fixed salaries than those who personally benefit from collected penalties, even when citizens’ material incentives are controlled for. Lower transparency reduces support, but only for salaried enforcers.

received valuable feedback.

1 Introduction

Enforcing the law is a primary function of governments, but regimes vary markedly in the extent to which citizens view their activities as legitimate. When authorities and the legal order are seen as legitimate, citizens may tend to obey the law even when the likelihood of punishment is slight (Tyler 2006). In contexts where authorities and the legal system lack legitimacy, levels of compliance may be lower and citizens may be less willing to associate themselves with the actions of law enforcement. In such settings, government officials may be met with indifference or hostility even by law-abiding citizens — witness widespread and longstanding African-American distrust of urban police forces (Tuch and Weitzer 1997; Weitzer and Tuch 2002) or Sicilian communities bound by the code of *Omerta* not to aid Rome in its fight against organized crime (Paoli 2003). In yet other places, government officials may be so loathed that they fear to tread the streets except in large numbers: for example, the Favelas outside of Rio de Janeiro and São Paulo (Caldeira 2000), the now-demolished Kowloon Walled City in Hong Kong (Girard and Lambot 1993), and the Five Points Neighborhood in 19th century Manhattan (Asbury 1928).

What institutions enhance or degrade the perceived legitimacy of government officials and their actions? The answer to this question can have large effects on societal welfare. Governments are tasked with monitoring citizen compliance with prescribed behavioral norms and punishing noncompliance where necessary. A principle widely acknowledged among law enforcement officials, regulators, and even military analysts is that in the absence of unlimited state capacity, these tasks can only be performed effectively if citizens generally acknowledge the legitimate authority of the officials performing them (Sampson, Raudenbush, and Earls 1997; U.S. Army and Marine Corps 2006).

The reasons for this sentiment are twofold: first, citizens who perceive a regime as legitimate may opt to comply with its precepts even when the probability that noncompliance would be detected is extremely low (Levi and Sacks 2009). Second, citizens may *assist* the officials from that regime who attempt to carry out enforcement against others — for example, by providing information to authorities about illegal activity or actively assisting officials in the apprehension of a criminal (Tyler and Fagan 2008) — thus enhancing the deterrent effect of enforcement. As such, legitimacy can reduce the need for enforcement and increase the efficacy of attempted enforcement, to the collective benefit of citizens.

Observational research on government performance has highlighted accountability, low levels of perceived corruption, the rule of law, and transparency as important correlates of citizen trust in government (Treisman 2007). However, establishing a causal link between legitimacy and these often-correlated in-

stitutional features remains challenging when using observational data.

We present the results of a public goods experiment designed to explore the institutional determinants of state legitimacy. We focus on two structural variables: the method by which enforcers are *compensated*, and the degree of *transparency* present in their operations.¹ In our scenario, “citizens” in a group decide whether or not to contribute to a public good. An authority external to the group observes these contribution decisions and then chooses whether to “target” a single citizen (who may in fact be innocent or guilty of failing to contribute) for enforcement. If the authority targets a given citizen, the *other* citizens must then choose whether to *help* the authority at a cost to themselves (increasing the probability that enforcement is successful); to *hinder* the authority at the same cost (decreasing that probability); or to do nothing. If enforcement is successful, the targeted citizen pays a penalty.

We follow many scholars in using public goods games as a metaphor for social norm compliance (Ledyard 1995), and add to an extensive literature begun by Fehr and Gaechter (2000) on punishment in public goods settings. The work most relevant to ours involves third-party enforcement (Fehr and Fischbacher 2004; Bernhard, Fehr, and Fischbacher 2006; Dickson, Gordon, and Huber 2009) rather than opportunities for decentralized sanctioning (Gëurek et. al. 2006; Nikiforakis 2010). A central novelty of our design is to incorporate citizen interventions that facilitate or undermine the external enforcer. This feature links our research to studies of second-order free-riding (Heckathorn 1989; Panchanathan and Boyd 2004), in which punishment (or, in our case, willingness to aid an external enforcer) is itself seen as a collective good subject to collective action problems. Our analysis of a (centralized) enforcer’s legitimacy is thus related to evolutionary accounts of how reciprocity norms, social learning, and pre-commitment affect the willingness to punish (Sigmund et. al. 2010; Perc 2012).

Our approach offers two key methodological advantages for studying legitimacy: first, because participants are *randomly assigned* to institutional environments, we are able to avoid selection problems and other obstacles to causal inference that complicate observational studies. Second, our approach uses a *behavioral* measure of legitimacy: namely, citizens’ costly decisions to assist or hinder the authority (as opposed to doing nothing). This approach captures the notion that legitimate authority entails not just support for, but also *obligation* or *duty* to the authority (Hart 1994). As such, we offer an alternative to the most common methods used in social psychology, political science, and law, which rely on survey measures of attitudes about institutional fairness or support for an institution (Caldeira and Gibson 1992;

¹A related literature considers the effects of how rules and leaders are chosen (e.g., elections versus exogenous assignment) on legitimacy (Dal Bo, Foster, and Putterman 2010; Baldassarri and Grossman 2011; Grossman and Baldassarri 2012; Ben-Yoav, Hollander, and Carnevale 1983). We study different determinants of legitimacy, and also use a different (behavioral) measure of legitimacy, described below.

Tyler 2006; Tyler and Fagan 2008).

A 2×2 design varies the institutional environment along two dimensions. The *compensation* dimension contrasts a salary condition, in which the authority is paid a fixed wage, with an appropriation condition, in which the authority is compensated on the basis of penalties collected. (In all our conditions, the authority benefits from the public good.) This dimension is meant to explore Weber’s (1978 [1922]) well-known precept that legitimate authority in the “legal-rational” sense depends on the absence of appropriative incentives. The *transparency* dimension contrasts full information environments, in which group members can directly observe whether or not a targeted individual actually contributed to the public good, with limited information settings, in which they cannot. This dimension is meant to explore the notion, dating at least to Bentham (1791), that government actions are legitimated in part via the transparency with which they are conducted.

A game-theoretic analysis helps to elucidate the actors’ strategic incentives under varying institutional environments and provides a benchmark against which to compare behavior in the lab. For the finitely-repeated games we employ in the laboratory, all equilibria in all treatments involve *no citizen intervention* in the aftermath of enforcement. The unique equilibrium for our salary treatments (under both full and limited information) involves no public goods provision and no enforcement by the authority; for our appropriation treatments (again, under both full and limited information), multiple equilibria exist, some involving positive enforcement and contribution levels.

In our experiment, consistent with equilibrium predictions, appropriator authorities under both full and limited information are significantly more resolute (willing to target non-contributors) and more predatory (willing to target contributors when all citizens contribute) than salaried authorities. However, citizen behavior at both the contribution and intervention stages departs significantly from game-theoretic expectations.

First, under full information, group members are substantially more likely to aid salaried authorities than authorities with appropriative incentives. We demonstrate that this difference is not driven by predatory authorities in the appropriation treatment, by citizens in the salary treatment compensating for weak authorities, or from unusually high public-spiritedness among citizens in the salary treatment. The data thus imply that the institutions themselves, above and beyond the differential behavior of authorities and citizens under those institutions, drive these differences (cf., Easton 1965). Crucially, we control for citizens’ material incentives when measuring these across-treatment differences in legitimacy, highlighting that legitimacy should be construed here as involving citizens’ intrinsic, non-material motivations to assist

authorities. We therefore demonstrate that these non-material motivations vary across institutional settings.

Second, we document several notable differences between the full and limited information treatments when authorities are salaried. Most strikingly, citizens contribute significantly less to the public good under limited information. We trace this difference to citizens’ reduced willingness to assist authorities when they are uncertain about a target’s culpability. This aloofness, in turn, undermines the deterrent effects of enforcement. Importantly, this finding strongly suggests that citizens do not trust salaried officials to behave justly in the absence of transparency, *even though those officials have no incentive to behave unjustly*. In the appropriation treatments, by contrast, contribution, enforcement, and citizen intervention behavior are nearly identical under full and limited information.

Our findings have several implications for efforts to refine positive accounts of government legitimacy since Weber. First, they demonstrate a previously undocumented tradeoff associated with employing high-powered incentives for public officials: while useful in motivating resolute enforcement, such incentives can degrade officials’ standing with citizens. This, in turn, can undermine the efficacy with which they perform key functions of governance. The mere fact that enforcers benefit directly from their enforcement efforts seems enough to diminish their legitimacy, even if these incentives do not lead them to behave badly. Second, our results suggest a novel *behavioral* justification for transparency: authorities with *motives* commonly known to be benign may nonetheless suffer a drop in support if citizens cannot directly observe that these authorities’ *actions* are in fact benign.

2 Background: Legitimacy, Incentives, and Transparency

While political legitimacy has a distinguished history in normative political thought (Peter 2010), efforts to quantify and measure legitimacy or describe its strategic (e.g., game-theoretic) underpinnings are less well developed (Hyde 1983). Max Weber (1978) [1922] provides the seminal analysis of legitimacy as a descriptive concept, defining legitimacy as a belief in or orientation to the validity of a social order — one that involves the recognition that the rules governing behavior within that order are binding on its subjects and “constitute a desirable model for him to imitate” (Weber 1947, 124).

We focus on two potential institutional sources of government legitimacy. The first is the means by which public officials (especially bureaucrats) are *compensated*. In describing “legal-rational” authority, Weber writes of “a complete absence of appropriation of his official position by the incumbent, and that

“[Staff] are remunerated by fixed salaries in money” (Weber 1978 [1922], 219, 220). For Weber, fixed compensation assures the objectivity of public officials, who can direct their actions toward neutral, prescribed ends rather than personal enrichment or advancement.

The most immediate objection to the Weberian model is that officials may be insufficiently motivated to exert effort performing their responsibilities absent some high-powered system of incentives (Becker and Stigler 1974; Harris and Raviv 1978). Indeed, departures from the Weberian ideal are ubiquitous throughout history: *tax farmers* (such as the *publicani* in ancient Rome, *pronoiai* under the Byzantine Empire; the *fermiers Généraux* in *ancien régime* France; and holders of medieval English fiefdoms) were granted authority to collect revenue on behalf of the state; excess revenue thus represented pure profit. Appropriative incentives might also be informal, as in environments where corrupt officials can extort bribes from citizens accused of offenses (Rose-Ackerman 1978). Quasi-appropriative behavior may also emerge from the career concerns of salaried officials: regulators or law enforcement officials may be rewarded based on the number of arrests, citations, or convictions they can claim credit for (Knowles, Persico, and Todd 2001; Gordon and Huber 2002; Huber 2007).

Our analysis of the relationship between official incentives and legitimacy is closely related to literature on the economics of agency, which identifies conditions under which high-powered incentives are undesirable (Holmstrom and Milgrom 1991; Dewatripont, Jewitt, and Tirole 1999; Benabou and Tirole 2006; Gailmard and Patty 2007; Wilson 1989); and to experimental research on reactions to perceived selfishness and/or unfairness (Lin-Healy and Small 2013; Fehr and Fischbacher 2002).

A second feature of government institutions long posited to affect their legitimacy is *transparency*. In a discussion of representative assemblies, Bentham (1791) advocates the “superintendence of the public” as a solution to a moral hazard problem between representatives and constituents. He also argues that publicity helps to secure the confidence of the people in their officials, arguing, “Suspicion always attaches to mystery. It thinks it sees a crime where it beholds an affectation of secrecy; and it is rarely deceived.” Contemporary scholars and nongovernmental organizations have stressed the salutary effects of transparency on government performance and accountability (Sen 1999; Besley and Burgess 2002; Ferraz and Finan 2008; Alt and Lowry 2010). Moreover, the *absence* of transparency is a defining feature of non-democratic regimes’ reliance on secret police forces to suppress dissent (Friedrich and Brzezinski 1965). Uncertainty about culpability may make untargeted citizens reluctant to “get involved” in enforcement relative to the full information environment; alternatively, it may place citizens in the thrall of the apparently arbitrary exercise of state power (Linz 2000).

3 The Experiment

3.1 Experimental Protocol

We conducted 14 experimental sessions at the NYU Center for Experimental Social Science lab to explore the dynamics of legitimacy. Each of the 225 subjects who participated took part in one session only. Subjects interacted anonymously via networked computers. The experiments were programmed and conducted using the software z-Tree (Fischbacher 1999). Participants signed up via a web-based recruitment system that draws on a large, pre-existing pool of potential undergraduate subjects. (Subjects were not recruited from the authors' courses, and did not receive course credit for participating.) After giving informed consent according to human subjects protocols, subjects received written instructions that were subsequently read aloud to promote understanding and induce common knowledge of the experimental scenario. No deception was employed, in accordance with the long-standing norms of the lab in which the experiment was carried out. Before beginning the experiment, subjects took an on-screen quiz that both measured and promoted understanding of the instructions. Subjects earned tokens, convertible into dollars at the end of the experiment (40 tokens = US\$1) in amounts determined by the outcomes of play. Subjects' overall payoffs in a given session were equal to the sum of payoffs from each of the 20 periods (converted into dollars), plus a US\$7 show-up fee.

At the beginning of each session, subjects were randomly assigned to a group of five people, of which one was randomly assigned as an authority ("Role A," in the neutral parlance of the experiment) while the others were assigned as citizens ("Role B").² Group and role assignments remained fixed over 20 periods of interaction. However, on authorities' screens, individual group members in role B were labeled with an "ID number" between 1 and 4 commonly known to be randomly reassigned every period.³

Each period consisted of one play of the following extensive form game:

1. Each citizen makes a binary choice to contribute (to a "common pot") or keep entire endowment of 20 tokens.
2. Authority observes contributions, decides whether to target one group member for enforcement (at cost of 2 tokens) or not. If not, game ends with citizens observing contribution decisions and payoffs

²Appendix A contains a sample set of instructions to subjects.

³Authorities and citizens interacted repeatedly over 20 periods; thus, enforcement decisions could have consequences extending beyond the particular period in which a given decision took place. Significantly, however, because group members' labels (ID numbers) were randomly reassigned every period (as in Fehr and Gaechter 2000), individual citizens could not establish reputations. This design choice better reflects the dynamics of enforcement within a large population, where individuals interact irregularly with authorities.

realized. Conditional on targeting:

3. Untargeted citizens observe (a) contribution decisions, whether targeting occurred, and target ID (full information treatments); or (b) whether targeting occurred (limited information treatments).
4. Untargeted citizens *help* or *hinder* authority (at cost of one token) or do nothing.
5. Authority and citizens observe helping/hindering and whether enforcement was successful; payoffs realized.

Appendix B provides a detailed description of the game and its equilibria.

The game is built around a standard linear public good game: across treatments, all citizens (and the authority) received 0.4 times the amount in the common pot (in addition to tokens kept). Thus, in the absence of enforcement each citizen had an incentive to keep his tokens (for a payoff of 20), but all would benefit if everyone chose to contribute (which would yield a payoff of 32).

Each citizen who intervened on behalf of the authority increased the probability that enforcement would be successful by $\frac{1}{6}$; each citizen who intervened on behalf of the targeted group member decreased this probability by $\frac{1}{6}$. Given a $\frac{1}{2}$ probability of successful enforcement in the absence of intervention, enforcement would be successful for certain if all three non-targeted group members were to intervene on behalf of the authority, while enforcement would be unsuccessful for certain if all of these group members were to intervene on behalf of the targeted citizen. In the event of successful enforcement, the penalty suffered by the targeted group member was 30 tokens.

The four treatments differed along two dimensions. The first dimension is the compensation scheme for the authority. In the salary treatments, the authority received a fixed per-period endowment of 20 tokens, but was not a claimant on any sanction successfully imposed. In the appropriation treatments, on the other hand, the authority received no fixed per-period endowment, but was awarded 30 tokens for each sanction successfully carried out. The second dimension concerned the information available to the citizen (“Role B”) players. In the full information setting, those players learned the contribution decisions of each of the other players in that round and which, if any, was targeted for enforcement. Thus, in addition to seeing their compatriots’ behavior in the period, they could observe whether a targeted individual had or had not contributed. By contrast, in the limited information setting, untargeted citizens players could observe only *that* a fellow citizen was targeted – but not whether the target or others had contributed – and only learned the total contribution level for the group at the end of the round.

Four experimental sessions (involving 60 subjects) were devoted to the *Salary/Full Information* treatment, while four other experimental sessions (also involving 60 subjects) were devoted to the *Appropriation/Full Information* treatment. Additionally, three sessions (involving 55 subjects) were devoted to *Salary/Limited Information*, and three sessions (50 subjects) to *Appropriation/Limited Information*.⁴

3.2 Game-Theoretic Benchmark

As a benchmark for our subsequent analysis, we examine non-perverse symmetric weak perfect Bayesian equilibria to the stage game. Weak perfect Bayesian equilibrium is defined conventionally (see Appendix B). Symmetry implies (a) all citizens play the same strategies; and (b) given more than one similarly-situated enforcement target (e.g., two non-contributors), the authority targets each with equal probability. Finally, non-perverse entails focusing only on equilibria in which, when indifferent but given the choice, the authority will target a non-contributor over a contributor (and so will only ever target a contributor if all citizens contributed).

Given our substantive focus, the critical result is that because citizen interventions are costly but offer no prospective benefit, in *all possible equilibria of all treatments*, citizens should never intervene to help or hinder the authority.⁵

In the salary treatments, enforcement is costly and provides no prospective benefit to the authority; thus salaried authorities should never attempt enforcement. Accordingly, citizens, fearing no sanction, should not contribute. Thus, (No Contributions, No Enforcement, No Intervention) is the unique equilibrium strategy profile.

In the appropriation treatments, the expected benefit of sanctioning to appropriator authorities (30 tokens $\times \frac{1}{2} = 15$ tokens) clearly exceeds the cost. Thus, appropriators should attempt enforcement, anticipating no resistance from untargeted citizens. Three equilibria are possible given our parameterization: In the first of these, there are no contributions; the authority targets one non-contributor; and group members choose not to intervene. In the second, all group members contribute to the public good; the authority targets one *contributor*; and again, group members choose not to intervene. And in the third, each citizen contributes with probability 0.78; the authority targets one group member (a non-contributor unless all group members contributed); and again, no intervention takes place. Note that these equilibria

⁴We achieved excellent balance across all comparisons made below on the pre-treatment covariates we measured: age, sex, and prior experience in the lab. The *minimum* p-value across 12 unpaired t-tests was 0.26; the mean was 0.60. See Supplemental Appendix C for a table of means and balance statistics for pre-treatment covariates.

⁵We note that this result, along with the other key benchmark results in this subsection, extend directly to our finitely-repeated laboratory scenario through application of backward induction.

do not depend on any distinction between full and limited information: at the point when this information is or is not revealed, citizens' only remaining decisions involve interventions, which are never sequentially rational.

The multiplicity of equilibria emerges due to strategic complementarities: if no other group members contribute to the public good, then the threat of sanction, if attempted, is spread sufficiently thin to make non-contribution a best response. On the other hand, if all other group members contribute to the public good, a given group member who does not contribute will be targeted for certain (given non-perverse enforcement), making the expected sanction sufficiently high to induce compliance.

4 Results

4.1 Authority and Citizen Behavior under Full Information

Under full information, citizens earned an average of \$0.65 per period in the salary treatment, and a statistically indistinguishable \$0.67 in the appropriation treatment. By contrast, authorities' earnings under full information did differ significantly across treatments: salaried authorities earned \$0.98 per period, versus \$0.72 for appropriator authorities.

4.1.1 Overview of Subject Behavior

Figure 1 presents a comprehensive view of authority and citizen decision making under full information. The data are presented in 24 panels – one per group. Within each panel, the solid line traces the overall level of public goods contributions (indexed by the left tick marks) over time. First-period contribution levels vary widely, between 1 and 4 in both compensation treatments. Within the salary treatment, 7 out of 12 groups settle into a pattern of generally high contributions (3 or 4 citizens contributing) in later periods, while the other 5 groups settle into a pattern of generally low contributions (0 or 1 citizens contributing). Within the appropriation treatment, a similar distribution of group tendencies is observed.

FIGURE 1 ABOUT HERE

Authorities' decisions to attempt enforcement are also depicted within these panels. Each small shape within the panel represents a decision to target a group member in a given period — triangles when a non-contributor has been targeted or squares when the authority's target contributed. There is considerable variation in the extent to which authorities attempted enforcement in both treatments. A key point to

note, which we will analyze in greater detail below, is the paucity of “unfair” attempted punishment in which the authority targets a contributor. In the salary treatment, the authority *never* targets a contributor in 10 out of 12 groups (and does so only a few times in each of the other two groups). In the appropriation treatment, the authority *never* targets a contributor in 8 out of the 12 groups (doing so only once in two groups, but frequently in two groups).

Citizens’ intervention decisions are also depicted within these panels. The height of each triangle or square indicates the net level of intervention chosen by the three non-targeted group members taken collectively (indexed by the right tick marks). This measure ranges from -3 (all untargeted citizens help the target) to +3 (all untargeted citizens help the authority). The data in the panels exhibits considerable variation in the extent to which intervention took place (and on behalf of whom).

Finally, the figures also display whether each instance of attempted enforcement was successful or unsuccessful. Solid shapes (triangles or squares) represent instances in which enforcement was successful, while hollow shapes represent unsuccessful enforcement. The relationship between opacity and height reflects the fact that citizen intervention affects the probability of success.

Each panel provides its own narrative, describing subjects’ behavior and containing hints about their motivations. In G10, for example, the salaried official never needed to deploy his enforcement authority, because full contribution was sustained. In G16, citizens tended to contribute, a predatory authority repeatedly attempted to improve his payoffs by targeting a contributor; and these attempts were generally thwarted as untargeted citizens frequently aided the target.

4.1.2 First-Order Incentives under Full Information

Apart from citizen intervention choices, the strategic incentives implied by our model offer three predictions about contributions and targeting. Analyzing these decisions helps validate our study design by showing that behavior in the lab comports, on average, with these theoretical predictions. First, we should rarely, if ever, see authorities engage in perverse targeting — attempting to sanction a contributor when at least one group member did not contribute. Indeed, of the 140 group-period observations in which at least one group member did not contribute but not all failed to contribute, in only seven was a contributor targeted: three out of 68 instances under the Salary/Full Information treatment, and four out of 72 instances under the Appropriation/Full Information treatment.

Second, we should observe that appropriator authorities target citizens more than salaried ones. In the absence of perverse targeting, more frequent targeting can come in the form of greater *resoluteness*

– a propensity to target non-contributors; or in the form of greater *predation* – a propensity to target contributors when *all* citizens contribute. Table 1 displays the frequency with which authorities engage in resolute and predatory targeting, excluding the small number of perverse cases.

TABLE 1 ABOUT HERE

The comparisons in the table strongly confirm our expectations (associated p -values are less than 0.001). For example, appropriator authorities behaved resolutely in 86% of the 105 group-period interactions with at least one non-contributor, but only in 62.5% of the 120 such cases under the salary treatment. The difference in predation across treatments is also stark: In 20% of the 131 interactions in which all citizens in the group contributed, appropriator authorities behaved in a predatory fashion; by contrast, predatory behavior occurred in just 1.7% of the 117 comparable cases for salaried authorities.

Third, we should observe evidence of *deterrence*: in particular, contribution levels should be higher where non-contributors are sanctioned at a high rate and contributors are sanctioned at a low rate. To test for the presence of this association, we created a *lagged deterrence index*: in a given group at period t , the index is calculated as the rate in the previous $t - 1$ periods at which non-contributors were successfully punished *minus* the rate at which contributors were successfully punished.⁶ Higher values of the index should correspond to greater levels of deterrence, and thus, higher contribution levels.⁷ Figure 2 plots contribution rates against the lagged deterrence index (by group-period), along with treatment-specific local polynomial smoothers and their associated 95 percent confidence intervals. As is evident in the figure, the data are consistent with deterrence in both of the full information treatments. Although we are careful to note that these relationships should not be interpreted as causal, they are consistent with our expectations about the study design.

FIGURE 2 ABOUT HERE

4.1.3 Citizen Contribution Rates

We turn to a comparison of citizen contribution rates across treatments. Overall, groups with a salaried official averaged 2.46 contributions per period, while their counterparts with an appropriator official averaged 2.83. These figures are not statistically distinguishable at the group level ($p = 0.58$, two-tailed).

⁶The measure is undefined for $t = 1$ or in later periods if for all $t - 1$ periods the group experienced contribution rates of either zero or 100 percent.

⁷Note that an observed relationship should not be interpreted causally, but merely as consistent with the first-order incentives.

Regression analyses of contribution rates disaggregated to the group-period level, with standard errors clustered at the group level, also show no significant difference in contribution rates across treatments (with and without period-specific indicators). Figure 3 displays smoothed polynomial estimates of average contribution levels per period (and their 95% confidence intervals) for the salary (black) and appropriation (white) treatments over all 20 periods of the experiment. In both treatments, contribution levels are sustained at a moderate level over time. A sequence of period-specific t-tests reveals that *in no period* can we reject the null hypothesis of equal contribution rates across treatments.

FIGURE 3 ABOUT HERE

Because our primary objective is to identify differences in citizen willingness to intervene that stem from considerations of legitimacy and not, for example, direct monetary incentives or the enforcer's behavior, the absence of differences in contribution rates (as well as the nearly identical earnings for the citizens in these two treatments, described above) is helpful from a research design perspective: we can be confident that observed differences in citizen intervention are not driven, for example, by differential response to prior levels of public goods provision, or by wealth effects. At the same time, it is worth noting that the absence of significant differences in contribution rates across treatments diverges from the equilibrium predictions above. The benchmark prediction for the salary treatment was *no contributions*; in the appropriation treatment, no contributions, full contribution, or an intermediate level of contribution were all consistent with equilibrium play. Thus, equilibrium predictions suggest (weakly) higher contribution levels in the appropriation treatment. The data, however, clearly indicate that this is not the case.

4.1.4 Citizen Intervention On Behalf of and Against Authorities

We now examine citizen intervention choices. At the most basic level, citizens clearly take into account the guilt or innocence of the authority's target. When an authority targeted a non-contributor, individual untargeted citizens intervened on behalf of the authority 30% of the time, and on behalf of the target just 8% of the time. (In the remaining 62%, citizens did not intervene). By contrast, when the authority targeted a contributor, individual untargeted citizens intervened on behalf of the authority just once in 105 opportunities and group members intervened on behalf of the targeted group member 61% of the time (and did nothing 38% of the time).

Citizens Help Salaried Authorities More. Under full information, there are substantial across-

treatment differences in citizens’ propensities to intervene on behalf of or against the authority. Conditional on attempted enforcement, net support for the authority (-3 to +3) was 0.89 in the salary treatment, as compared with -0.21 in the appropriation treatment ($p < 0.001$, two-tailed). In other words, on average, about one in three untargeted citizens chose to assist a salaried authority in her enforcement duties. Untargeted citizens intervened less often in the appropriation treatment – and more likely than not on behalf of the target.⁸

The left panel of Figure 4 provides a graphical summary of net assistance over time, at the group-period level: For all but the very early and very late periods, untargeted citizens are significantly more likely to intervene on behalf of the authority in the salary treatment (black) than in the appropriation treatment (white). Moreover, we observe statistically significant net *hindering* of appropriator authorities in periods ten to seventeen. By contrast, we observe statistically significant net *helping* of salaried authorities in all periods.

FIGURE 4 ABOUT HERE

The right panel depicts net assistance for the authority as a function of the average contribution rate of the group in prior periods. In groups with historically low levels of contribution, there is little intervention in either the salary or appropriations treatment, on behalf of either the authority or the target. By contrast, when historical contribution rates exceed 50% (i.e., two contributors per period on average), we see a marked (and highly significant) divergence between the treatments. In the salary treatments, net intervention jumps to over two. In the appropriation treatment, however, net intervention is negative for high levels of average contribution.

The figures indicate that the propensity to help or hinder varies as a function of variables other than the treatment. We therefore conducted a series of regression analyses to ensure that our results are not driven by a failure to control for additional observables. The dependent variable in these specifications is *Net Assistance to the Authority* (ranging from -3 to +3). Observations are groups in individual periods where enforcement was attempted. Coefficient estimates appear in the first two columns of Table 2. (Complete descriptions of the estimations carried out in generating this Table and Table 4 appear in Appendix C, where we also discuss potential threats to inference.)

⁸When aggregated up to the group level, the average net intervention for the salaried authority is 1.17, as compared with 0.43 for appropriator authorities. These differ from the disaggregated figures because enforcement occurs more frequently in some groups than in others. Two salary treatment groups that experience no enforcement drop out in the aggregated analysis. The p -value for the *unweighted* two-tailed comparison of group means is 0.13. Weighting groups by the frequency of enforcement, the p -value is approximately 0.01.

TABLE 2 ABOUT HERE

Specification (1) estimates the treatment effect controlling for period-specific indicators. Specification (2) adds three control variables: *lagged average group contributions*; *lagged average resoluteness* (coded as the proportion of prior periods in which any noncontributor was targeted, conditional on the presence of a noncontributor); and *lagged predation plus perverse targeting* (coded as the proportion of prior rounds in which any contributor was targeted). In both specifications, the difference between the appropriation and salary treatments remains highly statistically significant and comparable in magnitude to the results above. In specification (2), we find that the propensity to help or hinder the authority is strongly related to previous contribution levels. However, the resilience of the treatment effect in the presence of this control indicates that salaried authorities are not helped more in net because of different contribution rates under salaried and appropriative authorities (and recall that contribution levels were statistically indistinguishable across compensation institutions). Regarding our other control variables, we find no significant relationship between the authority’s past resoluteness and citizens’ propensities to intervene, but we find a strong negative relationship between propensity to assist the enforcer and past *bad* behavior by the enforcer.

Predatory Appropriators Do Not Drive the Result. Given this strong relationship, there may be some residual concern that appropriators’ relative illegitimacy may be due not to the appropriative *institution* itself but rather to the *behavior* of appropriators induced by that institution. A strategy more conservative than controlled regressions is simply to drop all group-period observations in which the group’s authority had *ever* targeted *any* contributor. The analysis in columns (3) and (4) of Table 2 accordingly includes only group-periods where the authority did not target a contributor in the current period and had not done so in any previous period. With this more conservative strategy, it remains true that appropriator authorities receive less assistance than salaried authorities (although the statistical significance of this difference is slightly diminished in the specification with minimal controls).

Citizens Are Not Substituting for Irresolute Enforcement. The above analysis allows us to calculate the probability that a given enforcement attempt is successful in each treatment. Recall that on average 0.89 untargeted citizens assist a salaried enforcer in net, while in the appropriation treatment, 0.21 untargeted citizens assist the target in net. Thus, conditional on targeting, an enforcement action would be successful $\frac{1}{2} + \frac{0.89}{6} = 64.8\%$ of the time in the salary treatment, and $\frac{1}{2} - \frac{0.21}{6} = 46.5\%$ of the time in the appropriation treatment. However, recall from Table 1 that appropriative authorities are more resolute, actually choosing to target non-contributors 85.7% of the time while salaried authorities do so

in 62.5% of opportunities. Multiplying these figures yields the probability of a successful enforcement outcome (conditional on the presence of at least one non-contributor): 40.5% in the salary treatment, and a nearly indistinguishable 39.9% in the appropriation treatment. The significant difference in citizen interventions almost perfectly cancels out the gap in enforcer resoluteness.

This suggests an alternative explanation: that citizens in the salary treatment, recognizing that authorities may be less resolute, may *substitute* for the authority’s resoluteness by intervening more often to maintain deterrence. The analysis in Table 2 controls for lagged resoluteness, mitigating against this concern. Nonetheless, we undertake two further analyses to evaluate this possibility. First, we carried out the column (3) and (4) specifications in Table 2 on the subsample of cases for which lagged resoluteness is 1 (i.e., the enforcer has always attempted enforcement given a lack of full contributions, and has never targeted a contributor). In this restricted subsample, we continue to find higher levels of assistance for salaried enforcers ($p < .05$, one-tailed tests); this cannot be due to substitution for observed past weak enforcement given the subsample composition.

Second, if citizens were substituting for irresolute enforcement, we would expect a negative relationship between the prior resoluteness of the authority and citizen intervention *within* the salary treatment. Moreover, we would expect resoluteness to have a larger effect in the salary treatment than in the appropriation treatment. Columns (5) and (6) of Table 2 contain separate effect estimates for the salary and appropriation treatments respectively. The findings are not supportive of the alternative explanation. The negative relationship between citizen intervention and resoluteness is actually estimated to be modestly stronger in the appropriation treatment than in the salary treatment (though one cannot reject a null hypothesis that the two effects are the same). The salary-treatment estimate is also not statistically distinguishable from zero (though its sign is negative).

Citizen Types Do Not Explain Differences In Legitimacy. Finally, we note that there is no significant difference in first-period contributions between the salary and appropriation treatments. This suggests no underlying difference in the baseline “public-spritedness” of subjects assigned to the two treatments.

4.2 Authority and Citizen Behavior under Limited Information

In this section we make three sets of comparisons: (1) between salaried and appropriator authorities under limited information; (2) between salaried authorities under full and limited information; and (3) between appropriator authorities under full and limited information.

Under limited information, group members earned \$0.54 per period on average in the salary treatment and \$0.62 under appropriation (a statistically significant difference). Authorities earned \$0.71 per period in the salary treatment, and a statistically indistinguishable \$0.70 under appropriation.

4.2.1 Overview of Subject Behavior: Limited Information Treatments

Figure 5 displays an overview of behavior under limited information. Under Salary/Limited Information, only one group sustains high contribution levels for the duration of the experiment. Other groups see initially high contributions fade, or consistently low contributions throughout. Contributions under Appropriation/Limited Information appear higher, with several groups sustaining higher contribution rates. There is also substantial heterogeneity in group members' willingness to help or hinder authorities.

FIGURE 5 ABOUT HERE

4.2.2 First-Order Incentives under Limited Information

We replicate our earlier tests of first-order incentives. Table 3 compares rates of resoluteness and predation in the salaried and appropriation treatments under limited information. As in the full information case, appropriator authorities enforce more frequently than their salaried counterparts. When all citizens contribute, appropriator authorities engage in predation 35% of the time under limited information, as compared to a lower 20% figure under full information.

TABLE 3 ABOUT HERE

The test for first-order deterrence effects requires elaboration. In the full information setting, our index of deterrence was an objective indicator capturing the frequencies with which noncontributors and contributors were targeted. In the limited information setting, citizens cannot generally be certain about the culpability of an enforcement target. As a consequence, deterrence is fundamentally subjective. However, citizens' beliefs should correlate strongly with the objective measure. For example, a citizen can infer predation with certainty if all citizens contributed and yet one was targeted, or resoluteness if none contributed but one was targeted; likewise, he can infer perverse targeting if he contributed and was targeted, upon learning at the end of the round that not all citizens contributed.

Figure 6 displays group-period level contributions as a function of the lagged deterrence index. The white lines represent locally smoothed polynomials (and their 95% confidence intervals) for Appropriation/Limited Information; black lines represent the smoother for Salary/Limited Information. A compari-

son to Figure 2 is revealing. First, the distribution of deterrence under limited information is considerably to the left of the distribution under full information (note that the horizontal axis now begins at -.25). Second, while we observe positive relationships between the deterrence index and contribution rates under both limited information treatments, those relationships are muted relative to their full information analogs.

FIGURE 6 ABOUT HERE

4.2.3 Contributions under Limited Information

Figure 7 displays smoothed polynomial estimates of average contribution rates, by treatment. The contrast between this limited information figure and its full information counterpart (Figure 3) is stark. Under full information, contribution rates were not appreciably different across compensation institutions, and exhibited no secular change in either treatment. By contrast, under limited information, the average contribution rate is *substantially* lower with salaried authorities than with appropriator authorities; moreover, average contributions decline over time in both treatments, this decline being more precipitous with salaried than with appropriator authorities.

FIGURE 7 ABOUT HERE

Does a decline in deterrence contribute to the drop in contributions under limited information when officials are salaried? If we assume that the deterrence index is an unbiased proxy for citizens' subjective beliefs, we can conduct a simple difference in means test comparing the deterrence index in the full and limited information salary treatments. Under full information, the mean deterrence index is 0.44. Under limited information, it is 0.17. The 0.26 difference is highly statistically significant ($p < 0.001$, two-tailed). Deterrence, of course, is driven by the behavior of both the authority (via targeting) and the citizens (via contributions and by helping and hindering). Comparing salaried officials in the limited and full information environments, there are no statistically significant differences in either resoluteness and predation rates, although resoluteness is modestly higher in the full information setting. This, in turn, implies that if the collapse of contribution rates under salaried limited information is driven by a concomitant collapse in deterrence, the source of the collapse must be a change in citizen intervention behavior, which affects the probability that attempted enforcement is successful.

4.3 Citizen Intervention On Behalf of and Against Authorities

Authority Compensation Method Has Reduced Impact under Limited Information. Figure 8 presents citizen willingness to help or hinder authorities under limited information, by method of authority compensation. The left panel displays group-period help/hinder propensities, along with local polynomial smoothers. The right panel displays propensity to help or hinder as a function of lagged average group contributions.

FIGURE 8 ABOUT HERE

The pictures differ markedly from their full information analogs (Fig. 4). The left panel indicates little, if any, difference between salary and appropriations under limited information: the confidence intervals of the smoothers substantially overlap for almost all periods. Moreover, the right panel lacks the clear separation across treatments under full information when previous contributions were high.

Since the analysis in the figures does not control for key covariates, we conducted a series of regression analyses for the limited information treatments. The results of these appear in the first two columns of Table 4. These estimates suggest that notwithstanding the graphical analysis above, it is, in fact, possible to replicate the finding from the full information treatments that citizens exhibit a higher propensity to help salaried than appropriator authorities. However, the finding is sensitive to specification: the estimate in the Column (1) specification, including only period-specific indicators, is not statistically significant ($p = 0.30$, two-tailed). The Column (2) estimate is about twice as large and is statistically significant at conventional levels ($p = 0.07$, two-tailed).

TABLE 4 ABOUT HERE

Limited Information Undermines Citizen Assistance of Salaried Officials. In columns (3) through (6) in Table 4, we analyze the effect of limited information *holding authority compensation constant*. Columns (3) and (4) restrict attention to the salary treatments. Limiting the information available to citizens is estimated to reduce net assistance for the authority by about .6 relative to the full information case. The effect is about 80% of the size of the difference between appropriation and salary treatments under full information (see column (2) of Table 2). In the column (3) specification, this result skirts conventional levels of statistical significance ($p = 0.107$) using a two-tailed test, though arguably a one-tailed test is appropriate given our *a priori* expectations about transparency's effects. In the column (4) specification, the result does reach conventional significance levels in a two-tailed test ($p = 0.08$).

By contrast, reduced transparency does not appear to affect citizens' willingness to assist an appropriator authority. As columns (5) and (6) show, appropriator authorities under limited information are estimated to garner slightly more support than appropriator authorities under full information, but these differences are modest and not statistically significant.

Limited Information Undermines Deterrence when Officials are Salaried. An analysis of the combined effects of authority and citizen behavior addresses why contributions are lower in the Salary/Limited Information treatment than in the other treatments. In the one-shot game, contribution is a best response for a citizen if she believes that the probability that she will be sanctioned if guilty (less the probability of sanctioning if innocent) exceeds $\frac{8}{30} \approx 0.27$. The deterrence index provides a useful proxy for this probability. In the Salary/Full Information treatment, 51% of group-periods have a lagged deterrence index above this threshold; under Appropriation/Full Information, 57% of group-periods do.

The corresponding figures are much lower under limited information: 29% under Appropriation/Limited Information, and merely 17% under Salary/Limited Information. This decline is larger in both absolute and proportional terms with salaried authorities.⁹

5 Discussion

Using a novel behavioral measure of legitimacy, we have explored the determinants of legitimacy by experimentally manipulating two institutional variables: how officials are compensated, and the transparency to citizens of authorities' decision making.

Our analysis suggests that an authority's legitimacy is directly affected by her institutional environment, even holding her behavior constant. Under full information, our findings confirm Weber's intuition that authorities' compensation schemes affect their (legal-rational) legitimacy. Weber's account, however, does not include a causal or even a particularly detailed argument about *why* appropriative incentives should undermine legitimacy.

One possibility, consistent with our results, concerns citizens' abilities to discern officials' true underlying motives based on their behavior. Imagine a world populated by two kinds of officials: venal and virtuous. If officials are salaried, venal ones will shirk unless the prospective benefits of deterring future noncompliance are substantial. By contrast, virtuous officials, whether through a strong commitment to

⁹The collapse of deterrence in the limited information salary treatment is compounded by the strategic complementarities in citizen contributions: the more (less) other citizens contribute, the stronger (weaker) the incentives to contribute, given limited authority capacity in targeting. As decreased assistance for the authority reduces the efficacy of sanctioning, incentives to contribute decline more steeply because of these complementarities.

the community's interests, or because of other intrinsic motivations, will be resolute. In this context, resolute enforcement serves as a costly signal of virtue. If officials have appropriative interests, however, both virtuous and venal types benefit materially from resoluteness. As such, citizens cannot discern an official's virtue through observing resolute enforcement. Given this, as well as evidence that benefiting personally from an action induces perceptions of selfishness (Lin-Healy and Small 2013), it would be understandable if citizens felt more obliged towards, and were more willing to help, salaried than appropriative authorities, even if authority behavior were the same under these different compensation schemes. Hermalin (1998) describes a model in which a leader can legitimate her actions through investing in a public good or undertaking other costly actions.¹⁰ However, Hermalin's leader signals private information about the marginal benefits of contributions, not private information about herself, as in our interpretation. Additionally, Hermalin's leader induces greater effort through exclusively instrumental motivations; by controlling for citizens' material incentives across treatments, our approach focuses on non-instrumental motivations for assisting authorities. In Levi (1988, 52-54), leaders encourage "quasi-voluntary compliance" by implementing a fair tax system and honoring their commitments. Her microfoundations focus on strategic complementarities in punishment: enforcement by a capacity-constrained authority is more effective when citizens are more willing to comply, as in our theoretical model. By identifying systematic variation in citizens' willingness to assist authorities, our experimental research complements her work by revealing an additional source of enhanced enforcement capacity.

Moving from full to limited information, levels of citizen support for appropriative authorities remain unaltered, while support for salaried officials collapses. This decreased support for salaried authorities appears at odds with rational accounts of how citizens might use information. Theoretically, salaried officials have no incentive ever to target a contributor: such an action yields no material benefit and could only serve to undermine the authority's deterrence powers in future periods. Empirically, authorities behave as expected, targeting contributors a mere three times in our Salary/Limited Information treatment. While citizens receive no direct feedback about the justice of enforcement attempts, they receive informative indirect feedback, for instance through observing that they themselves are never unfairly targeted, or that enforcers target nobody when everyone has contributed to the public good. As such, it might be expected that full vs. limited information would not affect citizens' expectations about enforcer behavior, and therefore also not affect citizens' intervention decisions. Instead, citizens in our experiments help salaried officials significantly less often under limited information.

¹⁰Ahlquist and Levi 2011 contains an extended discussion of related models.

Transparency appears instead to affect citizens' intervention behavior above and beyond the material value of the information that transparency entails. This suggests a novel *behavioral* rationale for transparency. If Bentham is correct and suspicion does attach to mystery, then it does so even when there is little actual mystery given the structure of the game. Lack of transparency undercuts legitimacy, and (indirectly) deterrence, not because enforcers are less resolute or more predatory, but because citizens do not give authority figures the benefit of the doubt about behaving in an appropriate fashion. Thus, benign motivation and behavior are not sufficient to engender legitimacy. Citizens appear also to need to verify the appropriateness of an enforcer's behavior. By contrast, in the appropriation treatments, enforcers' activities are already perceived as less legitimate even when they are transparent, so net deterrence differs little across informational contexts.

We conclude by noting that the results of our limited information treatments not only extend our full information findings, but also aid in their interpretation. That the gap in legitimacy between salary and appropriation under full information is not reproduced under limited information excludes, for example, the possibility that particular institutions induce unconditional senses of obligation to cooperate with authorities. Instead, citizens appear to internalize norms of cooperation in a way that is conditional, reciprocating publicly-observable fair conduct by officials. The flexibility of our framework will allow us to carry out additional treatments that could, in a similar way, further clarify the mechanisms through which institutions affect an authority's legitimacy. We plan to explore other facets of legitimacy's origins, including the role of elections, the role of social identities, and the dynamics of legitimacy where potential sources of authority compete. We also speculate that citizens' modes of psychological engagement with authorities may vary across cultural contexts; our framework could be adapted to measure cross-cultural variation in the dynamics of legitimacy using lab-in-the-field experiments.

References

- Ahlquist, John S., and Margaret Levi. 2011. "Leadership: What It Means, What It Does, and What We Want to Know About It." *Annual Review of Political Science* 14: 1-24.
- Alt, James E., and Robert C. Lowry. 2010 "Transparency and Accountability: Empirical Results for US States." *Journal of Theoretical Politics* 22: 379-406.
- Asbury, Herbert. 1928. *The Gangs of New York*. New York: Alfred A. Knopf.
- Baldassarri, Delia and Guy Grossman. 2011. "Centralized sanctioning and legitimate authority promote cooperation in humans." PNAS doi:10.1073/pnas.1105456108.
- Becker, Gary S., and George J. Stigler. 1974. "Law Enforcement, Malfeasance, and the Compensation of Enforcers." *Journal of Legal Studies* 3: 1-18.
- Ben-Yoav, Orly, Edwin P. Hollander, and Peter J. D. Carnevale. 1983. "Leader Legitimacy, Leader-Follower Interaction, and Followers' Ratings of the Leader." *The Journal of Social Psychology* 121 (1): 111-115.
- Benabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96: 1652-1678.
- Bernhard, Helen, Ernst Fehr, and Urs Fischbacher. 2006. "Group Affiliation and Altruistic Norm Enforcement." *American Economic Review* 96: 217-221.
- Besley, Timothy J., and Robin Burgess. 2002. "The Political Economy of Government Responsiveness: Theory and Evidence from India." *Quarterly Journal of Economics* 117: 1415-1451.
- Caldeira, Teresa P.R. 2000. *City of Walls: Crime, Segregation, and Citizenship in São Paulo*. Berkeley: University of California Press.
- Caldeira, Gregory A. and James L. Gibson. 1992. "The Etiology of Public Support for the Supreme Court." *American Journal of Political Science* 36(3 Aug): 635-664.
- Dal Bo, Pedro, Andrew Foster, and Louis Putterman. 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review* 100 (December): 2205-2229.
- Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies." *Review of Economic Studies* 66: 199-217.
- Dickson, Eric, Sanford C. Gordon, and Gregory A. Huber. 2009. "Enforcement and Compliance in an Uncertain World: An Experimental Investigation." *Journal of Politics* 71: 1357-1378.

- Easton, David. 1965. *A Systems Analysis of Political Life*. New York: Wiley.
- Fehr, Ernst, and Urs Fischbacher. 2002. "Why Social Preferences Matter – The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives." *The Economic Journal* 112 (Mar.): C1-C33.
- Fehr, Ernst, and Urs Fischbacher. 2004. "Third-Party Punishment and Social Norms." *Evolution and Human Behavior* 25: 63-87.
- Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90: 980-994.
- Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123: 703-45.
- Fischbacher, Urs. 1999. "Z-Tree Zurich Toolbox for Readymade Economic Experiments – Experiments' Manual." *University of Zurich Institute for Empirical Research in Economics Working Paper*, No. 21.
- Friedrich, Carl, and Zbigniew Brzezinski. 1965. *Totalitarian Dictatorship and Autocracy*. Cambridge: Harvard University Press.
- Gailmard, Sean, and John W. Patty. 2007. "Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Expertise." *American Journal of Political Science* 51: 873-889.
- Görek, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The Competitive Advantage of Sanctioning Institutions." *Science* 312: 108-111.
- Girard, Greg, and Ian Lambot. 1993. *City of Darkness: Life in Kowloon Walled City*. Watermark Press.
- Gordon, Sanford C., and Gregory A. Huber. 2002. "Citizen Oversight and the Electoral Incentives of Criminal Prosecutors." *American Journal of Political Science* 46: 334-351.
- Grossman, Guy and Delia Baldassarri. 2012. "The Impact of Elections on Cooperation: Evidence from a Lab-in-the-Field Experiment in Uganda." *American Journal of Political Science* 56 (4: October): 964-85.
- Harris, Milton, and Artur Raviv. 1978. "Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement." *American Economic Review* 68: 20-30.
- Hart, H.L.A. 1994. *The Concept of Law*, second edition. Oxford: Clarendon Press.
- Heckathorn, Douglas D. 1989. "Collective Action and the Second-Order Free-Rider Problem." *Rationality and Society* 1: 78-100.

- Hermalin, Benjamin E. 1998. "Toward an Economic Theory of Leadership: Leading by Example." *American Economic Review* 88: 1188-1206.
- Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7: 24-52.
- Huber, Gregory A. 2007. *The Craft of Bureaucratic Neutrality*. Cambridge University Press.
- Hyde, Alan. 1983. "The Concept of Legitimation in the Sociology of Law." *Wisconsin Law Review* 1983: 379-426
- Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109: 203-229.
- Ledyard, John. 1995. "Public Goods." In *The Handbook of Experimental Economics*, eds. John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press, 111-194.
- Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley, University of California Press.
- Levi, Margaret, and Audrey Sacks. 2009. "Legitimizing Beliefs: Sources and Indicators." *Regulation and Governance* 3:311-333.
- Lin-Healy, Fern and Deborah A. Small. 2013. "Nice Guys Finish Last and Guys in Last Are Nice: The Clash Between Doing Well and Doing Good." *Social Psychological and Personality Science*, published online 13 February 2013. <http://spp.sagepub.com/content/early/2013/02/13/1948550613476308>.
- Linz, Juan. 2000. *Totalitarian and Authoritarian Regimes*. Boulder: Lynne Rienner Publishers.
- Nikiforakis, Nikos. 2010. "Feedback, Punishment, and Cooperation in Public Goods Experiments." *Games and Economic Behavior* 68: 689-702.
- Panchanathan, Karthik, and Robert Boyd. 2004. "Indirect Reciprocity Can Stabilize Cooperation Without the Second-Order Free-Rider Problem." *Nature* 432: 499-502.
- Paoli, Letizia. 2008. *Mafia Brotherhoods: Organized Crime, Italian Style*. New York: Oxford University Press.
- Perc, Matjaz. 2012. "Sustainable Institutionalized Punishment Requires Elimination of Second-Order Free-Riders." *Nature Scientific Reports* 2: 344.
- Peter, Fabienne. 2010. "Political Legitimacy." In Zalta, Edward N., ed., *Stanford Encyclopedia of Philosophy* (Summer 2010). <http://plato.stanford.edu/archives/sum2010/entries/legitimacy/>
- Rose-Ackerman, Susan. 1978. *Corruption: A study in Political Economy*. New York: Academic Press.
- Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. Neighborhoods and Violent

Crime: A Multilevel Study of Collective Efficacy. *Science* 15: 918-924.

Sen, Amartya. 1999. *Development as Freedom*. Oxford: Oxford University Press.

Sigmund, Karl, et. al. 2010. "Social Learning Promotes Institutions for Governing the Commons." *Nature* 466: 861-863.

Treisman, Daniel. 2007. "What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?" *Annual Review of Political Science* 10: 211-244.

Tuch, Steven A., and Ronald Weitzer. 1997. "Racial differences in attitudes toward the police." *Public Opinion Quarterly* 61: 642-663.

Tyler, Tom R. 2006. *Why People Obey the Law*. Princeton: Princeton University Press.

Tyler, Tom R., and Jeffrey Fagan. 2008. "Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in their Communities?" *Ohio State Journal of Criminal Law* 6: 231-275.

U.S. Army and Marine Corps. 2006. *Counterinsurgency Field Manual*. U.S. Army Field Manual No. 3-24 and Marine Corps Warfighting Publication No. 3-33.5. Chicago: University of Chicago Press.

Weber, Max. 1947. *The Theory of Social and Economic Organization*. New York: Simon and Schuster.

Weber, Max. 1978 [1922]. *Economy and Society*. Berkeley: University of California Press.

Weitzer, Ronald, and Steven A. Tuch. 2002. "Perceptions of Racial Profiling: Race, Class, and Personal Experience." *Criminology* 40: 435-456.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

Table 1: The Effect of Authority Compensation Method on Resoluteness and Predation: Full Information Treatments

	Salary	Appropriation	Difference
Resoluteness	0.625 (0.044)	0.857 (0.034)	0.232 (0.056)
Predation	0.017 (0.012)	0.198 (0.035)	0.181 (0.037)

Standard errors in parentheses. Standard errors for difference in means calculated assuming unequal variances. Resoluteness is the rate at which the authority targets a non-contributor given that there is at least one in a group in that period, while predation is the rate at which the authority targets a contributor when all citizens contributed. Cases of perverse targeting excluded (see text).

Table 2: Predicting the Propensity to Help or Hinder Authorities, Full Information Treatments

	All					
	Observations (1)	(2)	Exclude groups with any targeting of contributors (3)	(4)	Salary Only (5)	Appropriation Only (6)
Appropriation Treatment	-1.004 (0.444)	-0.729 (0.284)	-0.614 (0.366)	-0.724 (0.270)		
Lagged Avg. Group Contributions		0.305 (0.123)		0.428 (0.115)	0.585 (0.113)	0.134 (0.098)
Lagged Avg. Resoluteness		-0.096 (0.309)		-0.239 (0.392)	-0.417 (0.472)	-0.623 (0.512)
Lagged Predatory or Perverse Targeting		-3.788 (0.441)			-13.085 (5.108)	-3.272 (0.476)
Period Indicators	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.262	0.506	0.258	0.463	0.566	0.495
N	200	188	160	148	73	115

Observations are group periods in which attempted enforcement took place. Dependent variable is Net Assistance to the Authority: the number of group members who assist the authority minus the number who hinder the authority (-3 to +3). OLS Coefficients with standard errors (clustered by group) in parentheses.

Table 3: The Effect of Authority Compensation Method on Resoluteness and Predation: Limited Information Treatments

	Salary	Appropriation	Difference
Resoluteness	0.551 (0.036)	0.870 (0.032)	0.319 (0.048)
Predation	0 (0)	0.355 (0.061)	0.355 (0.061)

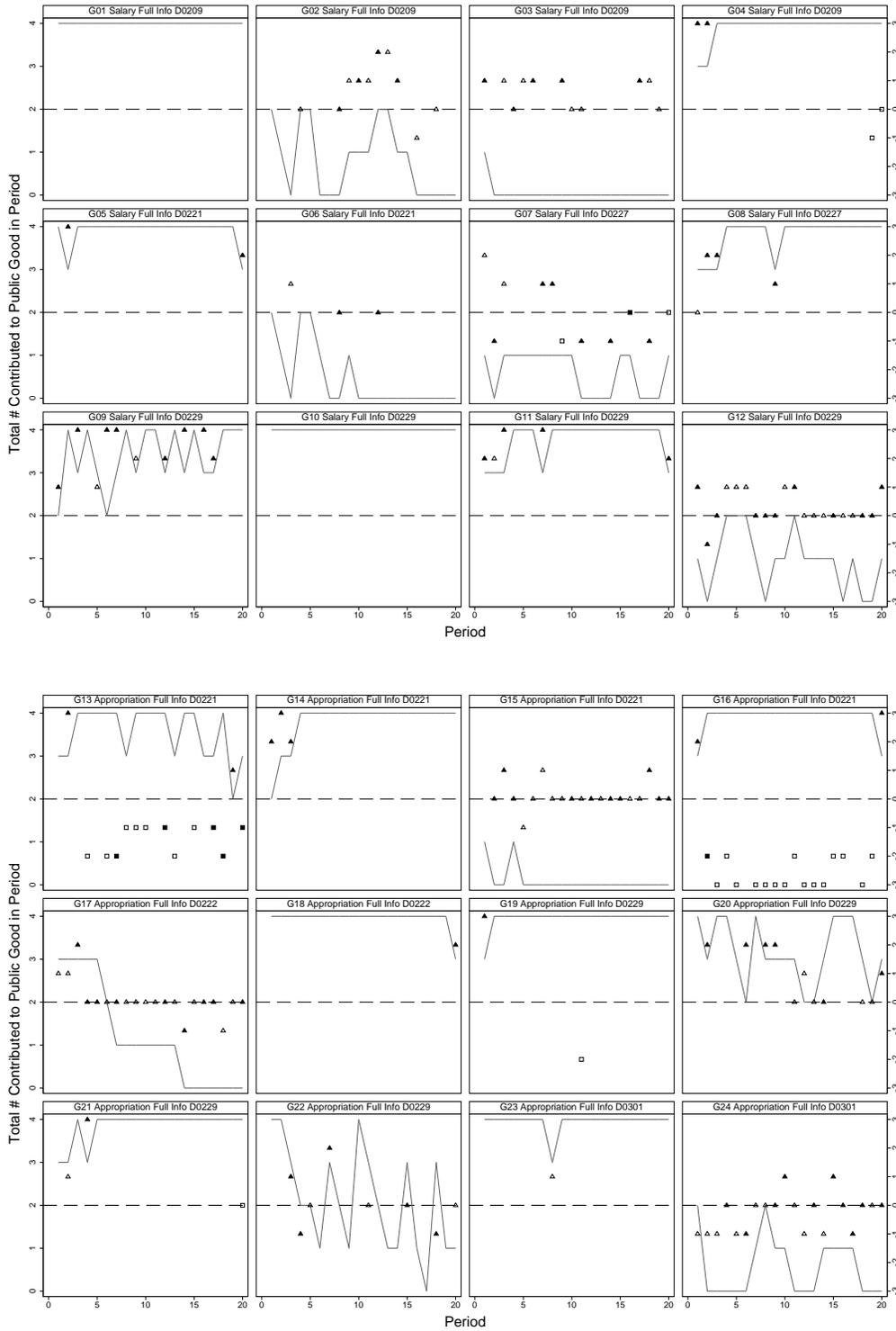
Analyses performed as in Table 1.

Table 4: Predicting the Propensity to Help or Hinder Authorities Conditional on Information and Authority Incentives

	Both Limited		Both Salary		Both Appropriation	
	Information Treatments (1)	(2)	Treatments (3)	(4)	Treatments (5)	(6)
Appropriation Treatment	-0.382 (0.359)	-0.788 (0.409)				
Limited Information			-0.643 (0.381)	-0.595 (0.322)	0.134 (0.423)	0.351 (0.474)
Lagged Avg. Group Contributions		0.398 (0.116)		0.440 (0.155)		0.159 (0.135)
Lagged Avg. Resoluteness		0.583 (0.535)		0.666 (0.530)		-0.106 (0.631)
Lagged Predatory or Perverse Targeting		-0.158 (0.769)		-4.336 (2.610)		-1.380 (0.847)
Period Indicators	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.080	0.164	0.120	0.290	0.095	0.164
N	257	244	192	179	265	253

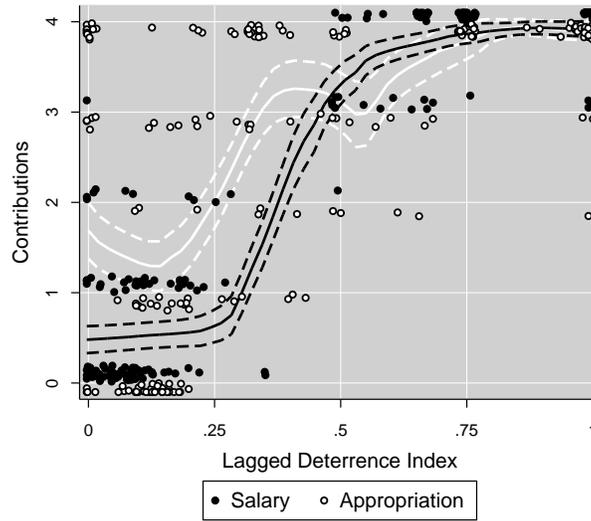
See caption for Table 2 for more information about the regressions.

Figure 1: Summary of Data from the Full Information Treatments



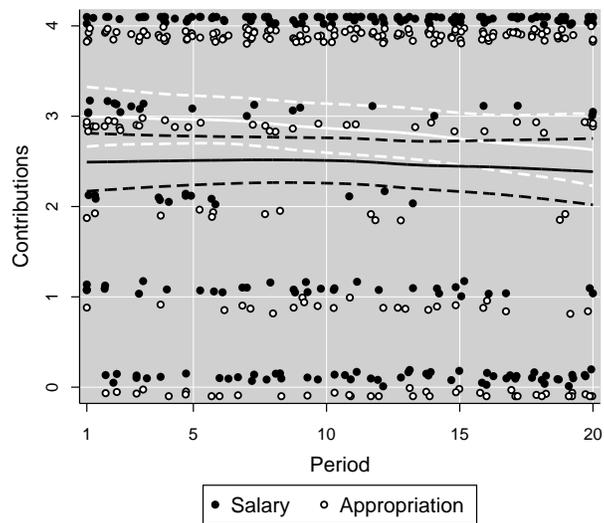
Each panel provides a comprehensive history of the group: the line represents contributions (as measured on the left axis); while symbols represent net helping/hindering (as measured on the right). Triangles denote a targeted noncontributor, and squares a targeted contributor. Hollow symbols denote enforcement failure, and solid symbols enforcement success.

Figure 2: Evidence of Deterrence Effects Under Full Information



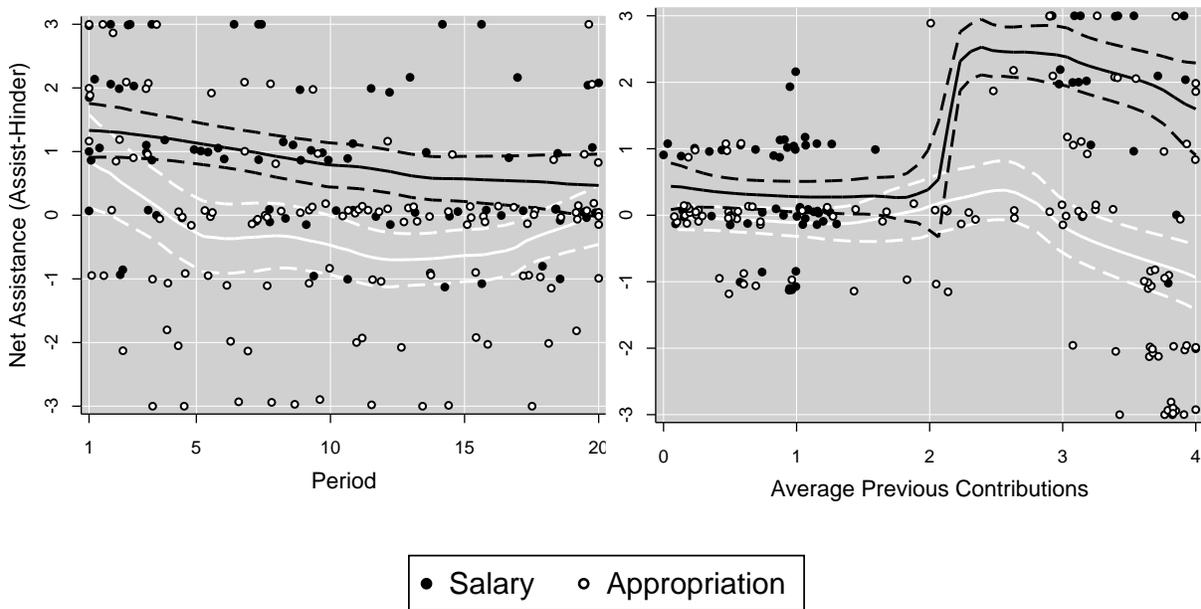
Each point corresponds to a group-period observation. Points are jittered and vertically shifted (by 0.1) from their true values for clarity. Solid curves denote treatment-specific local polynomial smoothers, and dashed curves their associated 95 percent confidence intervals. See text for an explanation of the deterrence index.

Figure 3: Contribution Behavior Over Time, Full Information Treatments



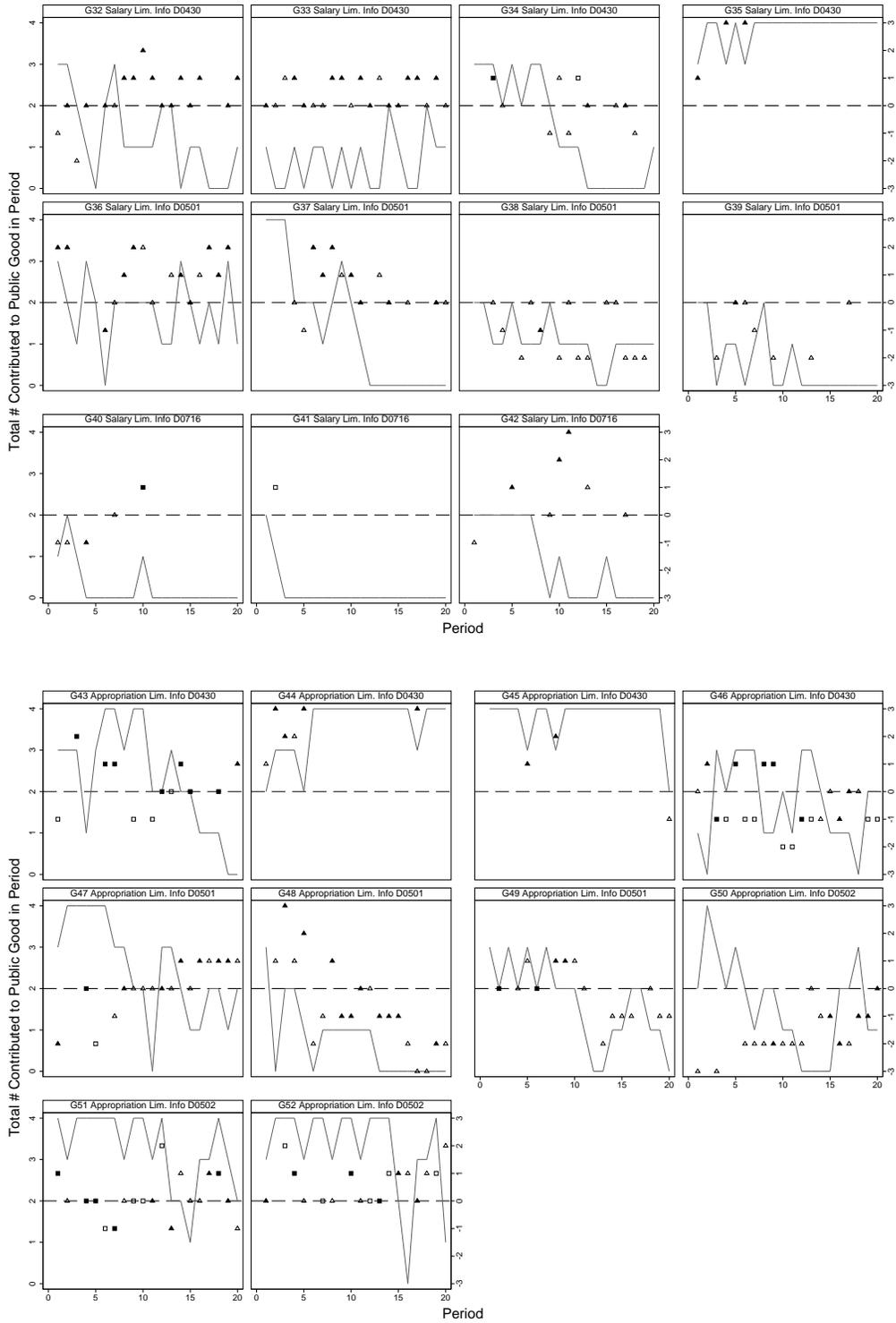
Each point corresponds to a group-period observation. Points are jittered and vertically shifted (by 0.1) from their true values for clarity. Solid curves denote treatment-specific local polynomial smoothers, and dashed curves their associated 95 percent confidence intervals.

Figure 4: Intervention on Behalf of or Against the Authority, Full Information Treatments



Each point corresponds to a group-period observation. Points are jittered and vertically shifted (by 0.1) from their true values for clarity. Solid curves denote treatment-specific local polynomial smoothers, and dashed curves their associated 95% confidence intervals. The left panel plots net citizen assistance over time. The right panel plots net assistance against the average previous contribution rates. Black denotes salary treatment, and white appropriation treatment.

Figure 5: Summary of Data from the Limited Information Treatments



See the Figure 1 caption for a key to this Figure.

Figure 6: Evidence of Deterrence Effects Under Limited Information

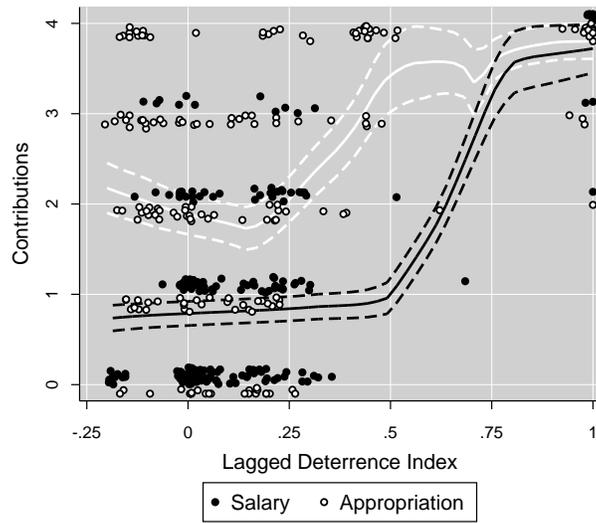


Figure constructed in same manner as Figure 2.

Figure 7: Contribution Behavior Over Time: Limited Information Treatments

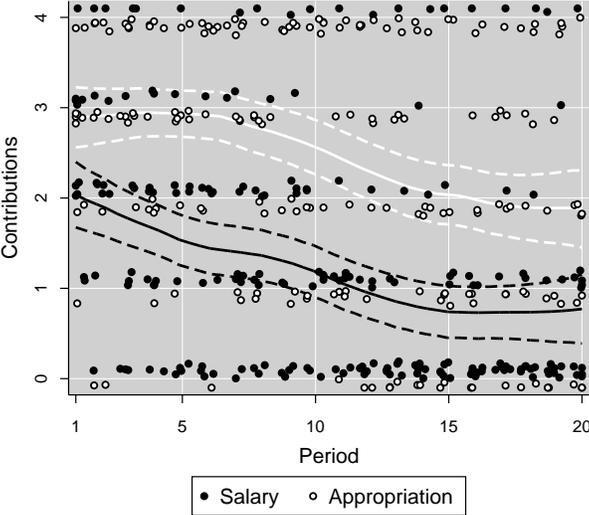


Figure constructed in same manner as Figure 3.

Figure 8: Intervention on Behalf of or Against the Authority, Limited Information Treatments

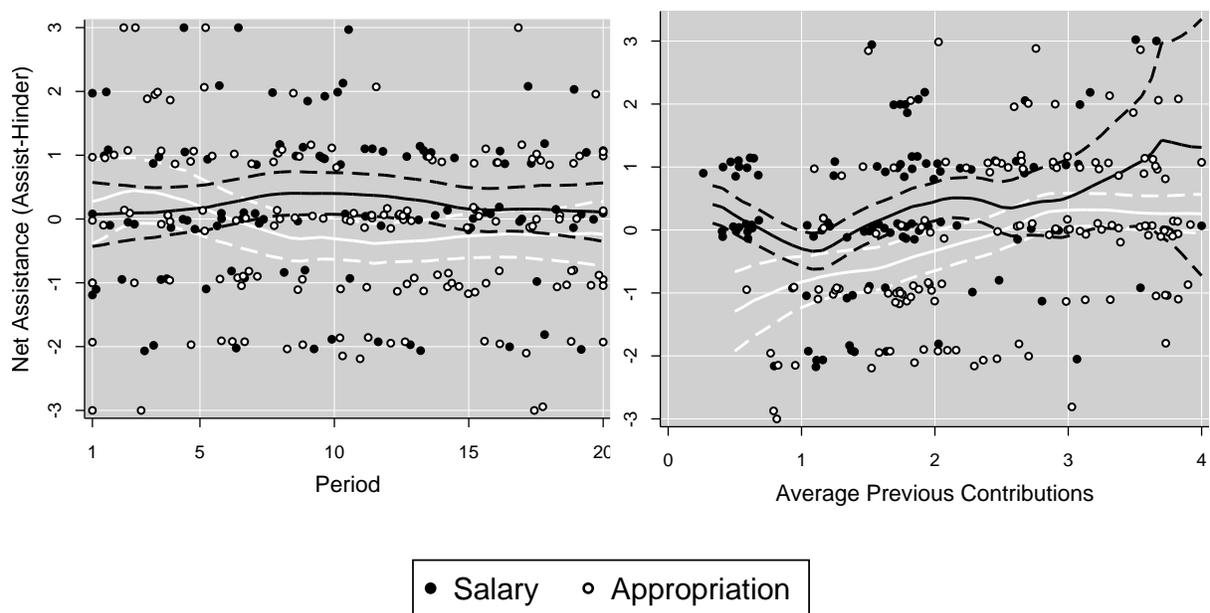


Figure constructed in same manner as Figure 4.