Identifying Legitimacy:

Experimental Evidence on Compliance with Authority

(Draft Version)

Eric S. Dickson
New York University

Sanford C. Gordon
New York University

Gregory A. Huber
Yale University

January 22, 2017

1 Introduction

This paper considers how an individual's decision to comply with rules and behavioral norms might be influenced by that person's perception of the legitimacy of the authority enforcing them; and the extent to which the choices of that authority, and the institutional context he or she occupies, enhances or diminishes those perceptions. The normative question of what constitutes legitimate authority has preoccupied political and legal philosophers at least since Plato, but concern regarding its empirical antecedents spans the social sciences, particularly psychology (e.g. Leventhal 1980, Tyler 1990) and sociology (e.g. Weber 1947, 1978 [1922]), but also political science (e.g. Easton 1965; Dickson, Gordon, and Huber 2015) and criminal justice (Bottoms and Tankebe 2012). The widespread interest in these issues is unsurprising: even in a world of limited state capacity, extensive citizen compliance with laws and social norms is still achievable if the citizens believe that compliance is, in a fundamental sense, "the right thing to do." By contrast, citizens who perceive authorities as illegitimate may be reluctant to assist them in the performance of their duties (Dickson, Gordon, and Huber 2015).

Beyond that, however, social scientists, philosophers, and other observers care about legitimate authority because they care about fairness, with the understanding that perceptions of unfairness can undermine legitimate authority, with deleterious effects for the social compact. A series of recent incidents around the United States involving aggressive police tactics, particularly against communities of color, have undermined trust in the police (USDOJ 2015) and brought these issues to the forefront of national political consciousness. An improved understanding of how authorities can effectively enforce the law on the one hand, while simultaneously treating citizens fairly and maintaining legitimacy on the other hand, could yield substantial benefits for citizen well-being. This paper takes one step toward engaging these issues, by attempting to isolate and measure the effects on citizen behavior of deterrence (through enforcement efficacy) and an authority's legitimacy (through her pursuit of procedurally fairer institutions) in a novel behavioral laboratory study. An understanding of the relative effects of these factors on citizen behavior is necessary when considering how best to formulate public policy to optimally addresses the tradeoffs inherent in fairly and effectively enforcing the law.

What is legitimacy? We proceed from Tyler (1997), who defines legitimacy as "judgment

by group members that they ought to voluntarily obey social rules and authorities irrespective of the likelihood of reward or punishment" (see also French and Raven 1959). In other words, legitimacy can be understood as a motivation to comply for irreducibly *non-material* reasons. Empirically identifying legitimacy and its behavioral and procedural antecedents is challenging, however, because of a fundamental problem of causal attribution in prior research that has claimed support for legitimacy-based accounts of compliance. The problem arises because procedures or choices that enhance an authority's legitimacy may simultaneously alter citizens' beliefs about the authority's capacity to bestow rewards and punishments. These material factors may, in turn, affect citizen behavior, quite apart from considerations of the authority's legitimacy.[1]

To sharpen the point somewhat, suppose a police officer behaves in a visibly "fair" manner, being scrupulous in her attention to detail and especially courteous to the suspects she interviews, and cultivating relationships with neighborhood citizens and business owners. If citizens on her beat obey the law more than they do a few blocks over, is it because they perceive her as legitimate, as existing literature often tends to interpret, or because they perceive her as competent and less error-prone? If her actions reduce the likelihood that she makes mistakes (arresting the innocent or failing to arrest the guilty), that will in turn improve deterrence, thereby improving compliance. But the relationship between her fair actions and citizens' compliance may also be driven by a positive affect toward the officer herself and the institutions she represents, which enhances citizens' intrinsic motivations to obey the law.

Or, consider the following example from the Department of Justice report on the Ferguson, Missouri police force. According to the report, "City, police, and court officials for years have worked in concert to maximize revenue at every stage of the enforcement process, beginning with how fines and fine enforcement processes are established." This, when combined with extant racial disparities in policing, severely degraded the citizenry's trust in the police. Now suppose that citizens in Ferguson, as a consequence of mistreatment, are less likely to comply with the law's edicts than those in a similar town with a less perverse law enforcement culture. A legitimacy-based account attributes this to the fact that citizens in Ferguson feel no psychological predisposition toward, or attachment to, the authority. But a deterrence-based account might attribute relative

[1]For a related critique of the analytical utility of the concept of legitimacy in the sociology of law, see Hyde 1983. Hyde suggests abandoning the concept altogether.

noncompliance to the fact that the police in Ferguson were less interested in appropriately punishing the guilty than they were in punishing as many people as possible, undermining material motives to comply.

Following the common practice of using public goods games as a metaphor for social norm compliance (Ledyard 1995), we derive a simple theoretical model that captures our understanding of how legitimacy concerns can motivate contributions to a public good over and above material incentives. In the model, centralized authorities vary in the extent to which they care about the fairness of an enforcement mechanism apart from the effect of fairness on the willingness of citizens to contribute to a common good, and can take a costly action to improve the mechanism's quality. citizens do not directly observe the degree of the authority's care, but can observe the costly action as well as the institution's resultant quality. citizens are motivated by material motivations – the return on the public good and the threat of punishment for failure to contribute. But they are additionally motivated to contribute by a "warm glow" (Andreoni 1990) that varies with two forms of legitimacy. The first, which we refer to as *institutional legitimacy*, corresponds to the warm glow associated with the fairness of the enforcement mechanism itself. The second, which we call *personal legitimacy*, corresponds to citizens' beliefs about the authority's degree of care. The model is close in spirit to the one described in Bénabou and Tirole (2003): in that setting, an authority with private information concerning her citizen's intrinsic motivation to exert effort is tasked with designing an incentive scheme for the citizen, but the nature of the incentive scheme is informative to the citizen about the authority's private information.[2]

Our theoretical analysis illustrates fundamental difficulties in disentangling deterrence and legitimacy-based motivations to contribute. These problems will hold under a very broad range of circumstances, potentially confounding all correlation-based studies purporting to document legitimacy effects even under random treatment assignment. All is not lost, however: the model also uncovers sufficient conditions under which identification of a personal legitimacy effect is feasible. These conditions correspond to situations in which costlier actions on the part of the authority correspond to "better" probability distributions over enforcement mechanisms. Under such conditions, informative equilibria exist in which authorities who care more about fairness undertake

[2]See also Hermalin (1998), in which a leader's "self-sacrifice" can credibly signal information about the returns to effort.

costlier actions, and citizens may observe and act based on variation in those actions while holding institutional quality constant.

We then describe the design, implementation, and analysis of two novel behavioral experiments that isolate a personal legitimacy effect induced by an authority's costly attempt to improve an enforcement mechanism's fairness. The experiments feature simple variants of the model's strategic environment in which, conditional on the authority's investment, the realized level of institutional quality is stochastic. Also, to prime the citizens' association of the authority with the punishment mechanism, the authority pre-commits to an enforcement strategy (e.g., "punish only those whom I observe not to contribute" or "punish everyone") following her costly investment in the quality of the institution.[3]

We find strong evidence of a personal legitimacy effect: specifically, the authority's mere attempt to implement a fair procedure increases the probability a citizen contributes by 10 to 12 percentage points. Thus, we provide direct behavioral evidence in support of the argument that legitimacy's normative core has positive behavioral implications. Furthermore, our approach allows us to compare the magnitude of these personal legitimacy effects to the legitimacy- and deterrence-based effects of institutional quality itself.

In the second experiment, we dig deeper into the psychological foundations of personal legitimacy by introducing an additional random element: after the authority chooses an investment and an enforcement strategy, but before citizens choose whether or not to contribute, a coin flip, observed by the citizens, determines whether the authority will materially benefit from the public good. Thus, we can assess whether personal legitimacy operates by activating a sense of reciprocity toward authorities among citizens, or through a warm glow independent of reciprocity. We find that the personal legitimacy effect is actually *larger* when authorities do not benefit materially from the public good than when they do. Thus, the personal legitimacy effect must be understood as a psychological motivation that exists over and above a feeling of debt to the authority.

[3]Insofar as "punish only those whom I observe not to contribute" is weakly dominant, this stage is omitted from the theoretical analysis. As described below, a large majority of subjects pursued this strategy.

2 A Model of Legitimate Authority

In the several decades since the seminal study of Thibaut and Walker (1975), scholars have explored the relationship between citizens' evaluations of an authority's commitment to procedural justice and their attitudes toward, and willingness to comply with, that authority (e.g., Lind and Tyler 1988; Tyler 1990; Tyler and Lind 1992; Tyler and Huo 2002; Paternoster et. al. 1997; Murphy 2004; Colquitt 2001). The channel through which this relationship is most often hypothesized to operate is legitimacy: deference to an authority and compliance with its edicts for non-material reasons. According to this account, procedurally fair institutions or behavior enhance an authority's legitimacy. Enhanced legitimacy, in turn, motivates citizens to comply. There are many ways in which an authority's procedures might be considered fair, and thus might enhance its legitimacy. In this study we focus on one in particular: accuracy (in the gathering of information and its application to the authority's decisions).[4]

To fix ideas, assume an authority and n citizens indexed by $i = 1, \ldots, n$, each with corresponding endowment normalized to one. Each citizen makes a binary choice $c_i \in \{0, 1\}$ to contribute the endowment to a public good or keep it for him or herself. Total contributions to the public good are given by $C = c_i + C_{-i}$, where $C_{-i} = \sum_{j \neq i} c_j$. We will refer to citizens who contribute as *compliant*, and those that withhold contribution as noncompliant. Citizens benefit from the return on the public good, $r(C)$, plus a "warm glow" $w(\cdot)$ from contributing, which we unpack in greater detail below.

The authority takes a costly action $a \in \mathbb{R}^+$ that affects the *quality* of an enforcement mechanism, $q := \mathbb{R}^+ \to (\frac{1}{2}, 1)$. We assume that $q(a)$ is nondecreasing in a. In the current context, quality refers to the probability that a sanction is correctly applied to compliant citizens and withheld from noncompliant citizens. The authority is characterized by a type, β, which is her own private information, and which scales the extent to which she cares about quality, *independent of its effect on contributions to the public good.* For example, the authority may loathe the prospect of "pun-

[4]Leventhal (1980) provides a comprehensive list that includes, in addition to accuracy, five additional criteria: (1) Consistency (across persons in space and time); (2) Bias suppression (absence of a personal stake on the part of the authority adjudicating disputes); Correctability (the opportunity to remedy previous mistakes, e.g., by appeal); (4) Representativeness (closely related to the notion of process control; that subjects believe that they have a say in the decision making apparatus); and (5) Ethicality (compatibility of the procedures with fundamental moral and ethical values). Research that experimentally manipulates dimensions apart from accuracy include Dickson, Gordon, and Huber 2015, Grossman and Baldassari 2012, Dal Bo, Foster, and Putterman 2011.

ishing the innocent" even holding constant the effect punishing the innocent has on contributions. The authority's utility is simply is simply $U_p = r(C) + \beta q(a) - a$.

The sequence of the game is as follows:

1. The authority observes her own type, β, and chooses an action a

2. citizens observe a and q, and simultaneously choose whether to contribute c_i.

The warm glow from contributing, $w(\gamma_i, q(a), \hat{\beta}(a))$, is positively affected by three factors. The first is an idiosyncratic motivation to contribute, γ_i, which is drawn from a continuous density with cdf $G(\cdot)$. The second is *institutional legitimacy*: the extent to which a higher quality institution engenders greater affinity toward it and hence toward the act of contribution. The third term is the authority's *personal legitimacy* in the eyes of the citizen: the citizen's posterior belief about the extent to which the leader cares about institutional quality. We indicate the dependence of $\hat{\beta}$ on a, noting that in the current context, a is a sufficient statistic for q and that in any pooling equilibrium, $\frac{\partial \hat{\beta}}{\partial a} = 0$.

In what follows, we will assume that $r(C) = \rho C$, where $\rho \in (\frac{1}{n}, 1)$ is the marginal per capita rate of return on the public good. The expected utility to the citizen of contributing and not contributing are, respectively,

$$\begin{aligned} E[U_i(c_i = 1)] &= \rho(1 + E[C_{-i}]) + w(\gamma_i, q(a), \hat{\beta}(a)) - (1 - q(a))s \\ E[U_i(c_i = 0)] &= 1 + \rho E[C_{-i}] - q(a)s, \end{aligned}$$

where s is an exogenous sanction level. Comparing these quantities, citizen i will contribute if and only if

$$w(\gamma_i, q(a), \hat{\beta}(a)) \geq (1 - \rho) - (2q(a) - 1)s. \tag{1}$$

Given monotonicity of $w(\cdot)$, the citizen's decision problem will be characterized by a cutpoint, γ^*, with the citizen contributing if and only if $\gamma_i \geq \gamma^*$, where γ^* is defined implicitly by (1) at equality. The probability that any individual citizen contributes is then given by $1 - G(\gamma^*)$.

The foregoing serves to elucidate the following fundamental identification issues inherent to the empirical study of legitimacy.

1. Deterrent and institutional legitimacy effects of quality are not generally separately identified.

From above, the probability a citizen contributes is monotonically decreasing in γ^*. Implicitly differentiating equation (1) at equality and rearranging yields

$$\frac{\partial \gamma^*}{\partial q} = -\left(2s + \frac{\partial w}{\partial q}\right) \Big/ \left(\frac{\partial w}{\partial \gamma}\right).$$

As is evident from the expression, an increase in q increases the potency of the material incentives to contribute (via the $2s$ term) while simultaneously increasing the positive affect toward contributing through the warm glow channel (via $\frac{\partial w}{\partial q}$).

This result has two immediate implications. First, it is a truism that large numbers of people comply with the law even if the chance of detected noncompliance is minuscule and the penalties slight. We would anticipate, even holding material incentives and idiosyncratic motivations constant, greater compliance under an institutional apparatus perceived by citizens as higher quality – via the institutional legitimacy channel.

The second implication concerns causal attribution: attempts to isolate a legitimacy affect by regressing, for example, for example, compliance (or stated willingness to comply) on institutional quality (or perceptions of quality) will be confounded by the extrinsic compliance incentive. Note that *randomization of the treatment does not fix this problem*: randomization, if feasible, would permit a valid causal estimate of the total effect of institutional quality, but not the effect attributable solely to institutional legitimacy.

In the social science literature, there are two standard empirical approaches to estimating the legitimating effect of procedurally fair institutions. The first is survey-based: query citizens who have had encounters with the law whether they were treated fairly, whether they perceive the institution as legitimate, and whether they intend to comply with the law in the future. This is the approach taken by Tyler (1990), who has documented a strong association between subjects' perceptions of the fairness of the legal process and the intention to comply with the law in the future. As other scholars have noted (e.g., van den Bos 2001), these studies are susceptible to potential problems of endogeneity and measurement error. The endogeneity problem arises because perceptions of fairness may be correlated with other unobserved individual- and institutional-level factors that also explain compliance. In the absence of a complete accounting of all differences

between treatment and control groups (those who experience more or less legitimate policing), the correlation between perceptions of legitimacy and compliance may be biased by factors correlated both with policing and rates of compliance. Measurement error may arise because citizens who report perceptions of police misconduct may misrepresent their lawfulness.

The second approach seeks to experimentally manipulate fairness: for example, Vermunt et. al. (1996) and van den Bos (2001) assign subjects to a condition in which an authority evaluates an exam by grading one or more answers. This research, however, faces a different measurement issue: the measured outcome is not compliance with the authority per se, but affect toward an institution. Our analysis suggests that even in the absence of this issue, this design remains subject to the causal attribution problem described here.

2. If $q(a)$ responds smoothly to changes in a, personal legitimacy effects induced by the authority's action are not separately identified from *either* deterrent or institutional legitimacy effects.

Implicitly differentiating (1) at equality and rearranging yields

$$\frac{\partial \gamma^*}{\partial a} = -\left[\left(2s + \frac{\partial w}{\partial q}\right)\frac{\partial q}{\partial a} + \left(\frac{\partial w}{\partial \hat{\beta}}\right)\left(\frac{\partial \hat{\beta}}{\partial a}\right)\right] \Big/ \left(\frac{\partial w}{\partial \gamma}\right)$$

The above expression shows three channels through which the action a propagates through the citizen's contribution choice. The first is the deterrent channel: higher a corresponds to higher q, which in turn corresponds to a greater motivation to contribute. The second is the institutional legitimacy channel: higher a corresponds to higher q, which corresponds to higher institutional legitimacy and a higher motivation to contribute. The third channel is the personal legitimacy channel: in any informative equilibrium, $\hat{\beta}(a_2) > \hat{\beta}(a_1)$ for any $a_2 > a_1$. Thus, a more costly action on the part of the authority results in a higher posterior belief on the part of the citizen about the authority's type, which, again, increases the motivation to contribute.

Consequently, regressing contributions on a measure of the authority's costly action will not separably identify a personal legitimacy effect unless the first two channels can be shut down. Note that for the same reasons as above, randomization does not fix this issue either. But the futility of random assignment is in some sense even starker here: the channel through which a operates on personal legitimacy is the posterior beliefs of the citizens about the authority's type in equilibrium. If a is randomly assigned, it will be uninformative with respect to type, and thus cannot affect

citizens' beliefs about the authority.

A necessary condition to recover an effect of the costly action a through the personal legitimacy channel (given the assumption of positive responsiveness of warm glow to q) is $\frac{\partial q}{\partial a} = 0$ at some q. In other words, the empirical challenge is to hold q constant while inducing variation in a. But this presents a dilemma. Note that the authority seeks to maximize her expected utility,

$$N\rho(1 - G(\gamma^*)) + \beta q(a) - a.$$

A necessary and sufficient condition for a separating equilibrium (in which authorities with higher β choose higher values of a) is that β and a are strategic complements. As is clear from this expression, this condition is violated if $\frac{\partial q}{\partial a} = 0$: authorities of higher type will lack the incentive to spend more to distinguish themselves from authorities of lower type.

The implication is that a condition that must hold to isolate personal legitimacy effects must be violated in order to induce the variation in the costly action needed to estimate those effects.

3. Identifying personal legitimacy effects is feasible if a induces lotteries over q with (partially) overlapping support.

Suppose that we sever the deterministic relationship between q and a in the following way: Let $\theta(q|a)$ represent the conditional density of q given the authority's action a, and suppose that for any a_1, a_2 with $a_2 > a_1$, (1) $\theta(q|a_2)$ is "better" than $\theta(q|a_1)$ in the sense of monotone likelihood ratio dominance; and (2) both $\theta(q|a_2)$ and $\theta(q|a_1)$ have common support in a nonempty interval (q_1, q_2). The authority seeks to maximize

$$E_{q|a}\left[N\rho(1 - G(\gamma^*)) + \beta q\right] - a.$$

Under these assumptions, β and a remain strategic complements, permitting separation on type. But the overlapping support of the conditional distributions implies that we can simultaneously observe different actions taken by the authority coincident with constant (or arbitrarily close) levels of institutional quality. These conditions permit us to separately detect a personal legitimacy effect.

The implication is that an experiment seeking to recover the personal legitimacy effect can do so by severing the deterministic linkage between the authority's behavior and the institutional

environment that the behavior affects. Note also that in preserving the belief-updating mechanism fundamental to the personal legitimacy channel, "treatment assignment" of different values of a to different citizens comes not through manipulation by the researcher, but via the natural variation in extent-of-care among the population of authorities.

3 The Baseline Experiment

3.1 Design

We designed an experiment to isolate how an authority's potentially legitimizing actions shape citizen compliance. In the experiment, an authority's chooses whether to try and improve the accuracy of the information that will be available to her when ultimately assigning punishments. The innovation of the design is to sever the deterministic relationship between the authority's decision and the actual institutional environment that eventually results, so that we can isolate the personal legitimacy effect as described above.

The experiment is embedded in a linear public goods game with a centralized authority: citizens choose whether or not to contribute to a public good and are subject to enforcement actions by the authority. The authority and the citizens share a common interest in maximizing contributions to the public good. Although individual citizens would prefer to withhold their contributions if all others contributed, citizens are nonetheless better off if all players contribute than if none do so. This approach follows a long tradition of using public goods games as a metaphor for social norm compliance (Ledyard 1995).

We conducted four experimental sessions at the NYU Center for Experimental Social Science lab and two sessions at the Yale Behavioral Research lab. Each of the 90 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 2007). At both institutions, participants signed up via a web-based recruitment system that draws on a large, pre-existing pool of potential student 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. 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 (30 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 period, subjects were randomly assigned to a group of five people, of which four were randomly assigned as citizens (Role A, in the neutral parlance of the experiment), and one as an authority (Role B). Group and role assignments were randomly reassigned after each period. In each period, individual group members in Role A were labeled with an ID number between 1 and 4, commonly known to be randomly assigned in each period.

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

1. The enforcer and each of the citizens receives an endowment of 20 tokens for the period.
2. The authority chooses to make a "Small Investment" or a "Big Investment" in information about each citizen's contribution to the public good. The authority pays a four token cost when choosing a "Big Investment," but pays no cost when making a "Small Investment."
3. Each group is randomly assigned to one of three different information institutions depending on the authority's choice. (1) Under "Low Accuracy Information," the information received by the authority about each individual citizen's contribution decision has a 40% error rate. If a specific citizen in fact kept (allocated) his tokens, the authority would correctly be informed that the citizen kept (allocated) his tokens with 60% probability, but would incorrectly be informed that he allocated (kept) his tokens with 40% probability. (2) Under "Medium Accuracy Information," the error rate is 25%. (3) Under "High Accuracy Information," the error rate is 10%.

   The authority's investment decision is linked probabilistically to an institution in the following way. If the authority chose a Small Investment, the group operates under "Low Accuracy Information" with 50% probability or under "Medium Accuracy Information" with 50% probability. If the authority instead chose a Big Investment, the group operates under "High Accuracy Information" with 50% probability or under "Medium Accuracy Information" with 50% probability.
4. After learning the quality of the information that she will receive (Low, Medium, or High), the enforcer commits to one of four binding enforcement strategies (which we label an enforcement rule) that will later be implemented, once she has received a signal about each citizen's contribution decision. The available enforcement strategies are (1) "Punish All": to punish all four citizens in the group, regardless of the signals received about citizen contribution decisions; (2) "PATS", short for "Punish According To Signal": to punish those citizens about whom a signal of "kept" is received, but not to punish citizens about whom a signal of "allocated" is received; (3) "anti-PATS": to punish those citizens about whom a signal of "allocated" is received, but not to punish citizens about whom a signal of "kept" is received;

and (4) "Punish None": not to punish all four citizens in the group, regardless of the signals received about citizen contribution decisions.

5. All citizens learn (1) the authority's investment decision (Small or Big Investment); (2) the actual accuracy of the information that will be available to the authority (Low, Medium, or High); and (3) the authority's chosen enforcement rule (Punish All, PATS, anti-PATS, or Punish None).

6. Each citizen simultaneously decides whether or not to contribute his or her entire endowment of 20 tokens to the public good, or to keep that endowment for him- or herself. The marginal per capita rate of return (MPCR) for contributions to the public good is 0.4; that is, an individual's decision to contribute to the public good yields a payoff of $0.4(20) = 8$ tokens for every citizen (and for the authority).

7. Signals to the authority about the contribution decision of each citizen are generated according to the relevant information institution (Low, Medium, or High) described above, with independent draws for each citizen. The pre-selected enforcement rule (Punish All, PATS, anti-PATS, or Punish None) is then automatically implemented, using the signals about citizen behavior sent to the authority when relevant. Any citizen who was punished has 24 tokens deducted from his payoffs; unpunished citizens suffered no deduction of tokens.

8. All players are informed about each citizen's actual contribution decision; the signal generated about each citizen's contribution decision; and which citizens were punished. Players are also informed about their payoffs for the period.

Before discussing identification, it is worth highlighting several features of citizens' instrumental incentives under this scenario. First, citizens' incentives to contribute to the public good are higher under PATS than under alternative enforcement strategies. Even in the presence of imperfect signals about citizen compliance, the authority's ability to deter non-contributions is highest when committing to a strategy of punishing those who appear to have been non-contributors and not punishing those who appear to have been contributors. Departures from PATS reduce citizens' material incentives to contribute, either by reducing the expected benefit of contributing (Punish All), increasing the expected benefit of not contributing (Punish None); or both (anti-PATS).[5]

Second, we selected the specific payoff parameters for our experiment to make citizens indifferent (in terms of their material incentives) between contributing and not contributing to the public good when an enforcer chooses PATS under a Medium Accuracy institution. To see this, note that a citizen who does not contribute under PATS/Medium Accuracy keeps his or her 20 token endowment but has 24 tokens deducted as a punishment with .75 probability, yielding an expected

[5]The choice of enforcement strategy may itself provide information to citizens regarding the authority's type, or her beliefs about the distribution of idiosyncratic compliance motivations among the subjects. Accordingly, all comparisons described below hold enforcement strategy fixed.

payoff of $20 - 0.75 \times 24 = 2$ tokens (based on the consequences of his own contribution decision; the payoffs he receives based on others' contributions are irrelevant to the present comparison). At the same time, a citizen who does contribute under PATS/Medium Accuracy receives a benefit of only $0.4 \times 20 = 8$ tokens from his own contribution, but has 24 tokens deducted as a punishment with only 0.25 probability, yielding an identical expected payoff of $8 - 0.25 \times 24 = 2$ tokens. As described next, our identification strategy focuses on comparisons within the Medium Accuracy institution given the PATS enforcement strategy (which we expected to be the most commonly chosen, because it maximizes deterrence incentives as noted above). The (material) indifference these parameter values induce is therefore useful because it maximizes our chances of detecting the effect of legitimacy on citizen motivations.[6]

Turning next to identification, we note that when citizens choose whether or not to contribute to the public good they are fully informed about all factors affecting their instrumental motivations to comply: the immediate payoff from keeping or contributing; the size of the potential penalty; the *realization* of the institution governing the quality of the authority's information; and the authority's pre-committed punishment strategy. By this time, they have also observed whether the authority took the potentially legitimating action – the "big investment" – that may inspire in citizens further non-instrumental motivations to comply. Critically, however, conditional on the realization of the accuracy level, *the authority's investment choice is materially irrelevant to the citizen*, and any institutional legitimacy afforded by the realization of the accuracy level is held constant. Consequently, any increase in contribution levels caused by the investment choice itself must arise through the personal legitimacy channel. Finally, because players are randomly assigned to new groups and roles at the end of each period and interact anonymously, they have no reason to condition their behavior in any stage of a given period on expectations of future actions or repeated interactions (e.g., beliefs about reciprocity).[7]

Given these features of the design, we are now in a position to make the experiment's instantiation of the identification strategy described in Section 2 explicit. Hold fixed the authority's enforcement strategy (e.g., PATS). Citizens will then be making their decisions in one of four circumstances, summarized in Table 1. Rows denote the authority's investment (Small or Big), while

[6] By contrast, in the presence of very strong material incentives to contribute or not to contribute, the effects of legitimacy on citizens' motivations to contribute would likely be masked.

[7] Below, we consider a psychological notion of reciprocity unrelated to repeated interactions.

Table 1: Identification of Legitimacy Effect in the Accuracy Experiment, Holding Punishment Strategy Constant

		Realized Accuracy Level		
		Low	Medium	High
Authority Investment Choice	Small	I. Low deterrence, Low inst. legitimacy Low personal legitimacy	II. Medium deterrence Medium inst. legitimacy Low personal legitimacy	N/A
	Big	N/A N/A	III. Medium deterrence Medium Inst. Legitimacy High personal legitimacy	IV. High deterrence High inst. legitimacy High personal legitimacy

columns denote the realized accuracy level. Conditional on the authority's choice of Low Investment, the citizen is randomly assigned into Cell I or II. Conditional on the authority's choice of Big Investment, the random assignment is to Cell III or IV.

A comparison of contribution levels of citizens in Cells I and II, or of citizens in Cells III or IV, yields a valid estimate of the total causal effect of accuracy stemming from the deterrence and institutional legitimacy channels. In each of these comparisons we hold the personal legitimacy of the authority's efforts fixed. Note that because we do not know the shape of the relationship between accuracy and institutional legitimacy, differences-in-differences will not yield an estimate of marginal deterrence fully purged of institutional legitimacy effects absent strong but untestable assumptions.

Our main focus, however, is on the comparison of contribution levels of citizens in Cells II and III, which yields a valid estimate of the personal legitimacy effect as manifested through the authority's effort to pursue more accurate information. This estimate is isolated to personal legitimacy because comparing these two cells holds fixed all material incentives channeled through the enforcement strategy and the realized accuracy level.

To be sure, there are other factors in the experiment that may also affect citizen behavior. Individuals' experiences during earlier rounds may affect their later play, and their behavior may generally evolve over the course of the game. While our design in which individuals are randomized into different groups and roles over time and then interact anonymously seeks to prevent most sources of repeat-play dynamics, we nonetheless take a variety of approaches to account for such dynamics in the analysis that follows.

3.2 Results

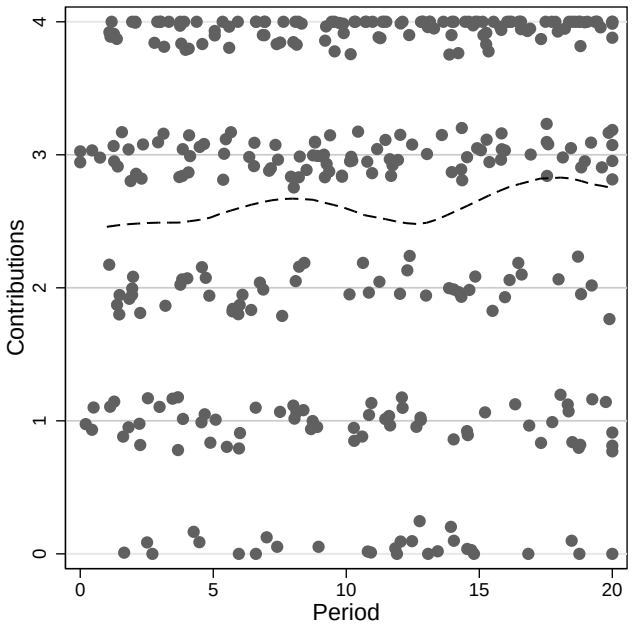
Summary of Authority and Citizen Behavior. The data consist of 360 group-period interactions (recall that in each period, players were randomly (re)assigned to groups and roles). Players assigned to the authority role chose the "big" investment 214 out of 360 periods, or 60%. This reflects a modest increase over time – in the first five periods the average rate was 50%, and in the last five it was 64%. Among players who were in the authority role more than once, 41% choose each investment level at least once. As a consequence of these choices, in 21% of all group-periods the authority received low accuracy information, in 53% medium accuracy information, and in 27% high accuracy information. 97% of players experienced all three accuracy levels while in the citizen role and the remaining 3% experienced two.

Those acting in the authority role overwhelmingly chose to punish according to the signals they received (PATS, 74% of the time), although a substantial minority chose to never punish (15%), and smaller proportions chose either to punish everyone (7%) or to punish only the apparently innocent (anti-PATS; 4%).[8] The PATS strategy is less common in the first 5 periods of play than afterwards: PATS is chosen by 67% of players in the first five rounds compared to 76% of the time in the remaining periods. Those who made the big investment in accuracy are more likely to choose PATS than those who made the small investment (87 versus 55% of the time), with those who made the small investment more likely to choose either to punish nobody (27 versus 7%) or to punish everyone (12 versus 3%). Because the investment decision affects the accuracy of the signals received by the authority, there is a similar relationship between accuracy levels and the enforcement rule.

Overall, citizens contributed their tokens to the public good 65% of the time, for an average group contribution rate of 2.6 out of 4. Unlike in standard public goods games with no enforcement (e.g., Fehr and Gachter 2000), contributions do not diminish over time. Figure 1 displays data on group-level contributions by period (data points are jittered for clarity), along with a local polynomial smoother. The average contribution rate rises slightly over time, from around 2.5 in the first five periods to 2.8 in the final five. 94% of players varied their contribution decisions, while the remaining players nearly evenly split between never and always contributing.

[8] Two players accounted for 47% of the cases of anti-PATS, while eight players chose the anti-PATS strategy only once.

Figure 1: Average Group Contributions by Period



Notes: Data jittered to enhance clarity of presentation. Dashed line is local polynomial smoother.

The decision to contribute likely reflects expectations about being punished depending on that choice. Overall, citizens who failed to contribute were punished 46% of the time. By contrast, citizens who did contribute were erroneously punished 24% of the time. Punishment "errors" – either failing to punish a non-contributor or punishing a contributor, occurred 35% of the time. These errors were less likely when the authority had high rather than medium or low accuracy information (17 versus 37 versus 52%). Focusing instead on the enforcement rule, punishment error rates were lowest under PATS (24%) compared to punishing everyone (35%), anti-PATS (72%), and never punishing (75%).

Confirming the importance of the importance of first-order institutional effects (via deterrence + institutional legitimacy), fixing the enforcement rule at PATS and the authority's investment at "big," group contributions are .5 units larger under high rather than medium accuracy ($p < .01$). Given PATS and a "small" investment by the authority, contributions within groups are 1.4 units larger under medium than low accuracy ($p < .01$).[9] Similarly, fixing information quality at medium,

[9]This analysis is based on an OLS regression predicting group contributions as a function of information levels among cases where the authority chose PATS. Note that better quality information should only affect deterrence when the enforcement rule appropriately responds to that information, which is the case only for PATS. Additional

group contributions are 1.8 units larger under PATS than the next best enforcement rule (punish all, $p < .01$), which has an outcome indistinguishable from either never punishing or anti-PATS.[10]

The Personal Legitimacy Effect. In order to isolate the effect of personal legitimacy concerns on behavior, we need to move beyond the simple descriptive analysis in the previous subsection to instead leverage the key design element of our behavioral game, which is the probabilistic link between the authority's investment decision and the realized accuracy level. Here and in our subsequent analysis we focus on citizen behavior as a function of the authority's investment decision, the realized accuracy level, and the authority's chosen enforcement rule. For this analysis we condition on the PATS enforcement rule because it is the dominant choice of authorities and because it is also the only enforcement rule for which expected deterrence improves with superior information. If the authority plays "always punish" or "never punish," then the information is ignored, while if the authority plays "anti-PATS," then more accurate information should *discourage* contribution; however, that strategy is so rarely used that it is not possible to systematically analyze its effects in any case.

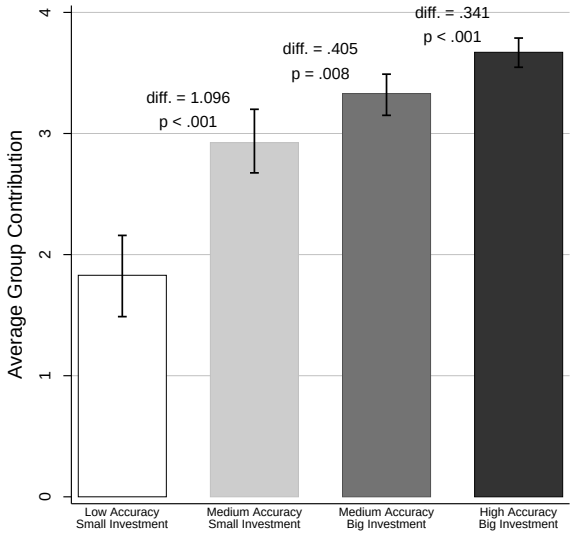
We first present the relevant data graphically in Figure 2, which plots group average contributions in cases where the authority chose PATS by the authority's investment decision (Small or Big) and the realized accuracy level (Low, Medium, or High). The leftmost white bar shows that when the authority choses the small investment and accuracy is randomly assigned to be low (rather than medium), 1.83 group members contributed on average (error bars represent 95% confidence intervals). By contrast, when the authority made the same choices but accuracy was randomly assigned to medium, the light gray bar shows 2.93 group members contributed. Thus, the randomly assigned increase in accuracy (corresponding to a comparison of cells I and II in Table 1) is associated with a 1.10 unit increase in contributions, a 27 percentage point increase ($p < 0.01$, two-tailed).

Next, suppose the authority chose big investment, so that random assignment was between medium and high accuracy. These realizations are represented by the two rightmost bars dark gray and black bars (corresponding to cells III and IV in Table 1). Switching from medium to high accuracy under big investment produces a .34 unit increase in group contributions (8.5 percentage

results in appendix table X.

[10]This analysis is based on an OLS regression predicting group contributions as a function of punishment strategy given medium accuracy signals.

Figure 2: Group-level Contributions by Investment Decision and Accuracy Levels Given the Authority Choses Punish According to Signal



Note: Averages with bootstrapped 95% confidence intervals.

points), from 3.33 to 3.67 out of four on average ($p < 0.01$, two-tailed). [11]

Finally, we now arrive at the main subject of our inquiry: the legitimating effect of the authority's costly *attempt* to improve the accuracy of the information available to her. Above, we demonstrated how to isolate this personal legitimacy effect from material and institutional legitimacy concerns by comparing subjects under medium accuracy and a fixed enforcement strategy between cases where the authority made the costly attempt to improve information and those where the authority did not. The middle two (light and dark gray) bars in Figure 2 permit this comparison for cases in which the authority adopted the PATS rule. Under medium accuracy and a small investment, the average group-level contribution is 2.93 out of four. Under medium accuracy and a big investment, the group-level contribution is 3.33 on average. The difference in means, about 0.41, is significant at $p < 0.01$ (two-tailed), and corresponds to an approximately ten percentage point increase in the contribution rate. Thus, a simple raw description of the data that leverages the design features of our experiment provides direct evidence of a personal legitimacy effect: authorities who attempt to improve the quality of information they receive, but who experience the

[11] This finding is robust at the individual level in a regression that includes period indicators, subject-specific fixed effects, and a measure of the subject's previous experience of contribution rates in prior rounds.

same realized level of quality as those who don't, induce greater contribution levels by their group members. Enhanced personal legitimacy therefore appears to be associated with real changes in costly citizen behavior.

Additional statistical analysis confirms this initial graphical presentation of our results. For this analysis, we shift from a focus on group-level outcomes to the decisions of individual players in the citizen role and model their contribution decisions (1=yes, 0=no) as a function of the treatment variable (the authority's big investment, 1=yes) and additional covariates. Data are restricted to cases in which the authority chose PATS and accuracy is medium. All specifications are OLS regressions with with standard errors clustered at the group-period level (i.e., a group of four citizens in a single group in a single period) because this is the level at which treatment is applied. Thus, our baseline specification is simply $Contribution_{i,t} = B_0 + B_1 \times BigInvestment_{i,t} + \varepsilon_{g,t}$ (where g indexes groups). This framework allows us to understand whether our results are affected by accounting for players' experiences earlier in the game, the period of play, or other factors.

Table 2 shows these results. Column (1) is a simple OLS specification predicting individual-level contributions that confirms the difference of means estimate of around 10% ($p < 0.01$, two-tailed).[12] Because there are four citizens in each group, this estimate is equivalent to the .41 estimate shown in Figure 2. Including period indicators in column (2) does not materially affect the estimated result, nor does accounting for each player's average experienced group contribution rate prior to this period (column 3). Players who have experienced groups with more contributions in the past are more likely to contribute, but accounting for this effect modestly increases our estimate of the legitimating effect of the authority's investment in the current period. In the column (4) specification, we also account for each player's experience of how frequently the authority in prior periods made the big investment. The estimated legitimacy effect remains positive and statistically significant.

In the column (5) specification, we restrict our analysis to players who had already served as the authority in at least one prior period (recall that subjects are randomly reassigned to new roles in each period), in case the absence of such prior experience meaningfully affects the way in which subjects understand the implications of the investment choice. In this specification we

[12]This specification clusters at the period-group level. Both unclustered and robust standard errors are smaller. If we instead cluster at the session level, standard errors are slightly larger, with $p = .06$, two-tailed.

Table 2: Estimated Effect of Authority's Investment Choice, Fixing Accuracy and Enforcement Rule

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Authority Big Investment	0.101	0.109	0.115	0.125	0.118	0.111	0.104	0.124
(1=yes, 0=small)	[0.038]***	[0.037]***	[0.037]***	[0.038]***	[0.048]**	[0.052]**	[0.039]***	[0.046]***
Average group contributions			0.072					0.078
prior to this period (0-4)			[0.030]**					[0.064]
Average big investment experience				-0.083				-0.047
prior to this period (0-1)				[0.081]				[0.144]
Constant	0.731	0.635	0.588	0.768	0.929	0.770	0.747	0.597
	[0.032]***	[0.063]***	[0.105]***	[0.098]***	[0.039]***	[0.063]***	[0.025]***	[0.139]***
Observations	560	560	540	540	404	338	447	428
R-squared	0.01	0.04	0.05	0.04	0.05	0.08	0.02	0.07
Number of unique subjects							65	65
Period fixed effects	N	Y	Y	Y	Y	Y	N	Y
Subject fixed effects	N	N	N	N	N	N	Y	Y

Dependent variable is player contribution decision (1=yes, 0=no).
OLS coefficients with group-period clustered standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.
Observations are individual contribution decisions given medium accuracy and PATS enforcement strategy.

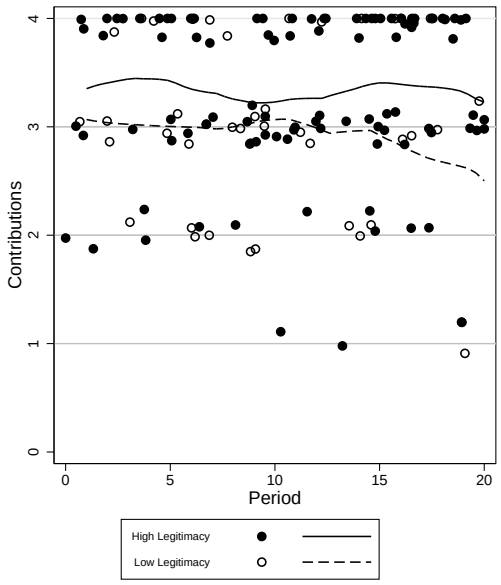
continue to find that individuals contribute more to the public good when the authority chooses a big investment. To reduce the possibility that our results are due to the behavior of subjects who did not understand the formal structure of the experiment, in column (6) we restrict our analysis to players who got at least 7 of the 8 "quiz" questions measuring subject comprehension correct prior to the beginning of the experiment.

Our most conservative analysis appears in columns (7) and (8). In these specifications, we include individual-level fixed effects and restricting the sample to players who experience medium accuracy and PATS in the citizen role following both a small and big investment by an authority. In the specification that does not include controls for period or past experience the estimate is .114 ($p < .01$), while it is .124 ($p < .01$) in the specification that includes those controls. Thus, even after accounting for each player's individual propensity to contribute, we still find evidence that the authority's legitimating choice of making the big investment directly increased player contributions to the public good.

Although our design seeks to rule out the possibility that the authority's investment choice affects any material concerns once we account for the realized accuracy level and the chosen enforcement rule – that is, that the choice affects only the legitimacy channel we highlight – one possible violation of this assumption could take place if citizens form expectations about the behavior of other citizens' likely behavior on the basis of the authority's investment. For example, in period 1 those citizens who see an authority choose a big investment may infer that other members of their group are more likely to be pro-social than would be the case if the authority had chosen a small investment. It is therefore encouraging that the specifications that account for average levels of previous group contributions and average prior authority behavior (i.e., columns (3) and (4)), which are direct measures of each citizen's prior experience and therefore proxies for each player's beliefs, do not reduce the estimated legitimacy effect.

As the authority's choice is likely to provide novel information about other citizen's proclivities, we would expect the authority's choice to be more influential in earlier rounds, when players have less experience with other players' contribution behavior. Each period provides three pieces of evidence about those tendencies (the contribution choices of one's fellow citizens), compared to only a single authority investment choice. For this reason, it is reassuring to find that the legitimacy effect is robust across subsets of the periods of the game. Figure 3 displays average contribution

Figure 3: The persistence of a legitimacy effect over time



Notes: Data jittered to enhance clarity of presentation.
Observations are group contributions given medium accuracy and PATS enforcement strategy.
Lines are local polynomial smoothers.

rates under medium accuracy and PATS, conditional on authority investment, over periods of play. Across all periods, average contributions are higher when the authority takes the legitimating action of the big investment.[13] More formally, partitioning the data into the first ten or last ten periods and re-estimating the specification shown in column (2) yields an estimated legitimacy effect of .097 in periods 1-10 ($p = 0.01$, two-tailed) and .130 in periods 11-20 ($p = 0.06$, two-tailed).

[13]The persistence of the legitimacy effect over time is also reassuring for another reason, in that it helps mitigate against explanations for our findings based on reciprocity across citizens. In standard public goods games without an enforcer, reciprocity has been shown to sustain cooperation when the shadow of the future looms large, but not in games (like ours) with anonymous interactions and random group reassignment. If reciprocity alone explained the pattern we observed, then we would expect it to be harder to sustain support for a big investment in later periods, when the game is about to end, than when there are many more rounds of play available. As the figure shows, however, we see no evidence that the legitimacy effect declines over time. Thus, it seems unlikely that this reciprocity account alone can explain the legitimacy result.

4 Exploring the Psychological Underpinnings of Personal Legitimacy

In the baseline experimental design, the authority benefits from citizen contributions to the public good. This design creates incentives for an authority to behave "appropriately" so as to maximize both her welfare and that of the citizens. However, because the authority also benefits from contributions to the public good, it is conceivable that citizen contributions to the public good are a way of "repaying" the authority for trying to do the right thing of improving citizen welfare. This form of reciprocity, which might persist as a way to encourage authorities to make the costly (probabilistic) investment in superior information, is analytically distinct from a pure warm glow accompanying favorable beliefs about the leader as modeled in Section 2.

To assess whether this citizen-authority reciprocity mechanism explains our main result, we conducted a follow-on experiment that breaks the linkage between the citizen's contribution decision and the authority's welfare. This design is similar to the original experiment, except that after the authority chooses an investment level, learns about the quality of information she will receive, and chooses an enforcement rule, an additional randomization takes place. With probability .5, the authority is assigned to the "public good compensation rule," in which the authority is paid as in the original experiment. Alternatively, the authority is assigned to the "flat fee" compensation rule, in which instead of being compensated on the basis of the public good, the authority is paid a flat fee for the period.

Importantly, this randomization takes place after the authority has made both of her decisions for the period but before the citizens have made their contribution decisions. Therefore, at the time the citizens are deciding whether to contribute to the public good, they both know all material factors that should affect their contribution decision (as in the original experiment) and whether the authority's welfare will be affected by their contributions. In the "public good compensation" condition, contributions are compatible with a citizen-authority reciprocity mechanism, while in the "flat fee" compensation condition each citizen's contribution has no effect on the authority's welfare and is therefore not affected by these reciprocity concerns.

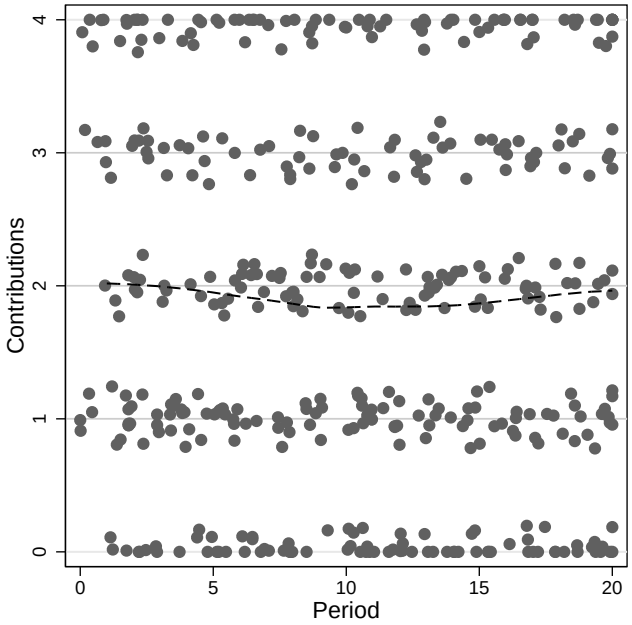
There are some other differences, which are detailed in the appendix. The most important is that for power reasons, we experiment with having two rather than three information accuracy

levels: specifically, conditional on the small investment by the authority, accuracy was high (20% error rate) with 25% probability and low (40% error rate) with 75% probability. Conditional on a big investment, accuracy was high with 75% probability and low with 25% probability. Thus, we are able to estimate personal legitimacy under both high and low accuracy, and under both compensation mechanisms.

We have data from 110 subjects gathered over seven sessions conducted at NYU and Yale. Briefly, authorities chose the "big" investment in 146 out of 440 periods, or 33% of the time. Correspondingly, accuracy was low in 58% of periods and high in the remaining 42%. Those acting in the authority role chose PATS 57% of the time, followed by never punish (22%), punish all (13%), and anti-PATS (9%). The lower rates of investment in this experiment may reflect the less stark effect of investments on information (there are only two levels of information, not 3); PATS is also chosen less frequently than in the first experiment, which may be explained by the correlation between the small investment and choosing never to punish. Overall, citizens contributed to the public good 48% of the time. While this rate is lower than in the first experiment, it is similarly stable over time (See 4). As with the first experiment, there is evidence that improved accuracy increases citizen contributions. Fixing the enforcement rule at PATS and accounting for the authority's investment decision, high accuracy is associated with a .37 unit increase ($p < .01$) in the rate of individual contributions.

Once again, our core analysis focuses on cases in which the authority plays PATS, fixes the quality of the information the authority receives (which may be high or low in this experiment), and estimates the effect of the authority's legitimating decision to invest in superior information. This analysis appears in Table 3. Column (1) estimates the pooled effect of the authority's investment decision, conditional on the realized information level, on citizens' contribution decisions. Per these results, the big investment is associated with a .07 ($p < .05$) increase in the probability each player contributes. In Column (2) we add indicators for whether the authority is paid from the public good and the interaction between that condition and the authority's investment decision. If the citizen-authority reciprocity pathway explains the results of the earlier experiment we would predict that the coefficient on Big Investment $\times$ Authority Enjoys Public Good to be positive because citizens would be more inclined to contribute when doing so could "pay back" the authority. We do not find this, however: The interaction between the compensation rule and the investment

Figure 4: Average Group Contributions by Period, Second Experiment



Notes: Data jittered to enhance clarity of presentation. Dashed line is local polynomial smoother.

decision is substantively small and statistically insignificant. However, the baseline effect of the investment decision (when the authority is not paid on the basis of the public good) – the estimate of the legitimacy effect that by design excludes this reciprocity mechanism – remains positive with a coefficient of .08 ($p = .06$, two-tailed test). This shows that even when reciprocity toward the authority is rule out as an explanation, we continue to find evidence that enhanced personal legitimacy operates through the warm glow channel and not from reciprocity motivations.

The remaining columns of the table assess the robustness of this result. As with our analysis of the first experiment, in column (3) we add period indicators, in column (4) we control for players' past experiences with group contribution rates, and in column (5) we account for past rates of authority investments. Including these control variables modestly diminishes the estimated effect of the investment in big information, to a minimum of .06 in column (4), and p-values (two-tailed) range from .10 to .12. Neither the effect of the authority being paid on the basis of the public good (which is consistently negative) nor the interaction between payment rule and investment decision are ever near statistically significant. Finally, the last two columns are player fixed-effects analyses. Including fixed effects diminishes the estimated legitimacy effect by about 50% in column (5), to

Table 3: Estimated Effect of Authority's Investment Choice, Fixing Accuracy and Enforcement Rule, Second Experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Authority Big Investment	0.066	0.076	0.063	0.058	0.066	0.034	0.012
(1=yes, 0=small)	[0.031]**	[0.041]*	[0.041]	[0.035]	[0.041]	[0.034]	[0.036]
Authority enjoys public good (1=yes)		-0.017	-0.034	-0.026	-0.023	-0.024	-0.031
		[0.042]	[0.040]	[0.036]	[0.040]	[0.034]	[0.036]
Big Investment × Authority		-0.019	0.000	0.013	0.012	-0.003	0.031
enjoys public good		[0.058]	[0.057]	[0.050]	[0.057]	[0.051]	[0.055]
High Accuracy (0-1)	0.370	0.369	0.387	0.403	0.390	0.386	0.413
	[0.032]***	[0.032]***	[0.032]***	[0.027]***	[0.032]***	[0.037]***	[0.039]***
Average group contributions				0.154			0.000
prior to this period (0-4)				[0.021]***			[0.034]
Average big investment experience					0.220		-0.031
prior to this period (0-1)					[0.075]***		[0.082]
Constant	0.444	0.452	0.511	0.113	0.410	0.462	0.500
	[0.025]***	[0.031]***	[0.055]***	[0.071]	[0.069]***	[0.023]***	[0.093]***
Observations	1000	1000	1000	960	960	1000	960
R-squared	0.18	0.18	0.19	0.25	0.21	0.24	0.26
Number of unique subjects						110	110
Period fixed effects	N	N	Y	Y	Y	N	Y
Subject fixed effects	N	N	N	N	N	Y	Y

Dependent variable is player contribution decision (1=yes, 0=no).
OLS coefficients with group-period clustered standard errors in brackets.
* significant at 10%; ** significant at 5%; *** significant at 1%.
Observations are individual contribution decisions given PATS enforcement strategy.

.03 (not significant), while also including controls for period of play and past group experience diminishes it further to around .01. While these estimates are small, we note that in neither case do we observe a positive interaction between the authority being paid from the public good and her investment decision, supporting our main conclusion that the legitimacy effect from the first experiment is not driven by the citizen-authority reciprocity pathway.[14]

5 Discussion

This paper makes several contributions to the study of legitimate authority and its relationship with procedural fairness. Methodologically, we demonstrate the challenge of isolating legitimacy from material motivations for compliance. The innovation of our experiment is to sever this relationship by making the legitimating *action* of the authority probabilistically, rather than deterministically, related to the materially-relevant *consequences* of that action.

At the same time, we are cognizant that the stylized environment in which we are able to implement this design may depart from the "real world" in a number of respects. To underscore this point, consider two of the potentially artificial features of the experiments described above: that the authority can pre-commit to an enforcement rule, and that the authority has capacity to punish all of the citizens in her group. The first of these features is particularly important because it means that each citizen in a group is operating with a common understanding of the environment of material incentives: they know the realized level of accuracy and the mapping between their behavior and the likelihood of punishment. In reality, of course, authorities tend not to pre-commit to enforcement strategies. In the absence of pre-commitment, an entirely different signaling account may operate, in which the citizens make inferences about the likely enforcement strategy from the (ex ante) procedural investment. In such an environment, it might be quite challenging to disentangle the non-instrumental effect of the legitimating action from the instrumental consequences of the information the action conveys to the citizens.

Likewise, the assumption that the authority possesses the capacity to target all citizens is clearly

[14]Given the baseline high rates of contribution under high accuracy and PATS, about 86%, it may be that the effect of legitimacy is obscured by a lack of systematic variation in contribution behavior given high accuracy. If we repeat the fixed effects analysis from column (5) but focus only on cases of low accuracy, the estimated coefficient on big investment is .10 (S.E.=.07) and the interaction of big investment and authority compensation rule remains negative and insignificant.

unrealistic given the significant capacity constraints that affect nearly every state apparatus. In the context of the experiment, this is a valuable feature because it allows us to eliminate the strategic complementarities that arise in most enforcement environments. If an authority's capacity is limited, in many circumstances there are multiple equilibria in which expectations about others' behavior come into play. If a citizen expects all other citizens to comply, non-compliance carries greater risks, because the probability of being targeted by the authority is higher; as such, the citizen's incentives to comply are heightened. If instead a citizen expects no or few other citizens to comply, non-compliance is less risky, because the probability of being targeted by the authority is reduced, weakening incentives for compliance. In an environment with strategic complementaries such as these, authorities' legitimating actions could affect citizen behavior by coordinating expectations about others' behaviors, thereby affecting equilibrium selection. Because such complementarities are excluded in our design, however, this alternative mechanism is beyond the scope of this study.

Additionally, our study departs from the corpus of social psychological research on procedural fairness in its use of an incentivized laboratory environment in which subjects receive financial compensation for their performance in a game. The value of this setup is twofold. First, it permits us to establish a benchmark of rational behavior based on purely material motivations against which to compare actual behavior. With this benchmark in hand, we can more definitively attribute observed differences in contribution behavior to specific psychological motivations. Second, because subjects benefit or suffer materially from their actions and those of other subjects, our experimental environment approximates the sorts of compliance choices that individuals must make in their day-to-day lives.

Our novel empirical approach also allows us to make progress in understanding the psychological origins of institutional legitimacy. Specifically, we show that leaders who attempt to obtain procedurally fairer institutions, operationalized here as those that are more accurate, have more legitimacy. Enhanced legitimacy, in turn, increases citizens compliance, and those effects are beyond those that arise due to changes in instrumental motivations. Note that it is not realized procedures that we use to generate variation in legitimacy. Rather, we show that an authority's willingness to seek a fairer procedure, independent of whether that procedure comes into being, enhances her her personal legitimacy and alters citizen behavior. A necessary step for future work is to un-

dertake additional research using different designs to understand whether the realization of fairer institutions also generates concomitant improvements in compliance via a legitimacy mechanism, although the methodological hurdles we identify here imply these efforts may be difficult.

Finally, there are numerous dimensions of procedural fairness aside from accuracy. For example, the authority may or may not take actions that appear biased against a member of an out-group. While the experiment described here focuses on accuracy, this design is sufficiently flexible to accommodate any number of other conceptions of procedural fairness. A task for future research is to estimate the legitimating effects of these different notions of fairness, and to see whether they are driven by the presence of the fairer procedure or the authority's costly investment in it.

References

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal* 100: 464-477.

Baldassari, Delia and Guy Grossman. 2011. "Centralized sanctioning and legitimate authority promote cooperation in humans." PNAS doi:10.1073/pnas.1105456108.

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.

Bénabou, Roland, and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." *American Economic Review* 70: 489-520.

Bernhard, Helen, Ernst Fehr, and Urs Fischbacher. 2006. "Group Affiliation and Altruistic Norm Enforcement." *American Economic Review* 96: 217-221.

Bottoms, Anthony, and Justice Tankebe 2012. "Beyond Procedural Justice: A Dialogic Approach to Legitimacy in Criminal Justice." *Journal of Criminal Law and Criminology* 102(1): 119-170.

Colquitt, Jason A. 2001. "On the Dimensionality of Organizational Justice: A Construct Validation of a Measure." *Journal of Applied Psychology* 86(3): 386-400.

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.

Dickson, Eric S., Sanford C. Gordon, and Gregory A. Huber. 2015. "Institutional Sources of Legitimate Authority." *American Journal of Political Science* 59(1): 109-127.

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 G$\acute{}$achter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90: 980-994.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10(2): 171-178.

French, John R., and Bertram Raven 1959. "The Bases of Social Power." In Cartright, D., and A. Zander, eds., *Group Dynamics: Research and Theory.* Ann Arbor, MI: University of Michigan Press, 150-167.

Friedrich, Carl, and Zbiegniew Brzezinski. 1965. *Totalitarian Dictatorship and Autocracy.* Cambridge: Harvard University Press.

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): 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.

Hermalin, Benjamin E. 1998. "Toward an Economic Theory of Leadership: Leading by Example." *American Economic Review* 88: 1188-1206.

Hyde, Alan. 1983. "The Concept of Legitimation in the Sociology of Law. *Wisconsin Law Review* 1983: 379-426

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.

Leventhal, Gerald S. 1980. "What Should Be Done with Equity Theory?" In Gergen, Kenneth J., Martin S. Greenberg, and Richard H. Willis, eds. *Social Exchange.* Boston, MA: Springer: 27-55.

Lind, E. Allan, and Tom R. Tyler. 1988. *The Social Psychology of Procedural Justice.* New York: Plenum Press.

Murphy, Kristina. 2004. "Regulating More Effectively: The Relationship Between Procedural Justice, Legitimacy, and Tax Non-Compliance." *Journal of Law and Society* 32(4): 562-589.

Paternoster, Raymond, Robert Brame, Ronet Bachman, and Lawrence W. Sherman. 1997. "Do Fair Procedures Matter? The Effect of Procedural Justice on Spouse Assault." *Law and Society Review* 31:163-204.

Thibaut, John W., and Laurens Walker. 1975. *Procedural Justice: A Psychological Analysis.*. Hillsdale, NJ: L. Erlbaum Associates.

Tyler, Tom R. 1990. *Why People Obey the Law.* New Haven, CT: Yale University Press.

Tyler, Tom R. 1997. "The Psychology of Legitimacy: A Relational Perspective on Voluntary Deference to Authorities." *Personality and Social Psychology Review* 1(4): 323-345.

Tyler, Tom R., and E. Allan Lind. 1992. "A Relational Model of Authority in Groups." *Advances in Experimental Social Psychology* 25: 115-192.

Tyler, Tom R., and Yuen J. Ho. 2002. *Trust in the Law: Encouraging Public Cooperation with the Police and Courts.* New York, Russell Sage Foundation.

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.

United States Department of Justice, Civil Rights Division. "Investigation of the Ferguson Police Department." March 14, 2015.

van den Bos, Kees. 2001. "Uncertainty Management: The Influence of Uncertainty Salience on Reactions to Perceived Procedural Fairness." *Journal of Personality and Social Psychology* 80(6): 931-941.

Vermunt, Riel., Arjaan Wit, Kees van den Bos, and E. Allan Lind. "The Effects of Unfair Procedure on Negative Affect and Protest." *Social Justice Research* 9(2): 109-119.

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.